63-1-1

RADC-TR 62-205

June, 1962

STUDIES IN LIE DETECTION

Computer Feasibility Considerations

Joseph F. Kubis

FORDHAM UNIVERSITY
New York, N. Y.

AF 30(602)-2270

Project No. 5534

Task No. 553401

Prepared

for

Rome Air Development Center
Air Force Systems Command
United States Air Force
Griffiss Air Force Base
New York

RADC-TR 62-205 June, 1962

STUDIES IN LIE DETECTION

Computer Feasibility Considerations

Joseph F. Kubis

FORDHAM UNIVERSITY
New York, N. Y.

AF 30(602)-2270

Project No. 5534

Prepared

for

Rome Air Development Center
Air Force Systems Command
United States Air Force
Griffiss Air Force Base
New York

TABLE OF CONTENTS

Chapter		Page
	LIST OF TABLES	iii
	ACKNOWLEDGMENTS	vi
	SYNOPSIS	vii
1.	THE LIE DETECTOR: Feasibility for Computer Analysis	1
2.	METHODOLOGY: Simulated Theft Experiment	5
3.	RESULTS: Simulated Theft Experiment	34
4.	DENIAL-OF-ACTUAL-CRIME EXPERIMENT	79
5.	DENIAL-OF-CLASSIFIED-INFORMATION EXPERIMENT	87
6.	COUNTERMEASURES EXPERIMENT	93
7.	DISCUSSION, CONCLUSIONS, AND RECOMMENDATIONS	106
	REFERENCES	115
	APPENDICES	
	A INSTRUCTIONS--Simulated Theft Experiment	116
	B INSTRUCTIONS--Denial-of-Actual-Crime-Experiment	132
	C INSTRUCTIONS--Denial-of-Classified-Information Experiment	136
	D INSTRUCTIONS--Countermeasures Experiment	139
	E RECORDING FORMS	144
	F PHOTOGRAPHS	147

LIST OF TABLES

Table		Page
1.	Distribution of Subjects by Class Year	28
2.	Classification of Incomplete and Discarded Records	32
3.	Accuracy of Examiners and Raters for Both Sections of the Simulated Theft Experiment	36
4.	Chi-Square Tables and Chi-Square Values (X^2) for Examiners' Data	38
5.	Consistency between Examiners and Raters in Terms of Frequency of Complete Agreements	39
6.	The Innocent-Thief (I $\rightarrow$ T) and the Thief-Innocent (T $\rightarrow$ I) Errors for the Simulated Theft Experiment	40
7.	Percent Accuracy of Examiners in Each of their Dual Roles (as Examiners and as Raters).	41
8.	Three Types of Percent Agreement Scores for Each Examiner.	44
9.	Examiner Accuracy Percent Scores in Both Sections of the Simulated Theft Experiment.	47
10.	Percent Agreement Scores for three Rating Conditions in Both Sections of the Simulated Theft Experiment	48
11.	T-I (Thief-Innocent) Discriminant Scores for the Physiological Indices, the Total, and the Discriminant Function (Typical Set of Data).	50
12.	The Relative Accuracies (as Percentages) of the Several Scores for the Thief-Innocent, Lookout-Innocent, and Thief-Lookout Classifications (Data from Entire Experiment).	52
13.	The Relative Accuracies (as Percentages) of the Several Discriminant Scores for the First and Second Sections of the Experiment	56

Table | Page

14. Relative Ratios of the Thief-Innocent Discriminant Weights for All Examiners and Raters in the Second Section of the Experiment 58

15. Relative Ratios of the Thief-Innocent Discriminant Weights of three Examiners Who Served as Raters for Several Other Examiners 59

16. Chi-Square Values for Critical Words and Numbers Denoting the Efficiency of Discrimination for the T-I, L-I and T-L Pairings of Roles in the First and Second Sections of the Experiment 62

17. Correlation Coefficients between Respiratory, Plethysmographic and Psychogalvanic Discriminant Scores for the Thief, Lookout and Innocent Suspect Categories in the First and Second Sections of the Simulated Theft Experiment . . 64

18. "Average" Correlation Coefficients Expressing Degree of Objectivity Among Raters in Evaluating the Physiological Responses of Thief, Lookout, and Innocent Suspect for the First and Second Sections of the Experiment . . . 67

19. Averages of Confidence Ratings for Correct and Incorrect Identifications of Roles of Subjects for First and Second Sections of Experiment 68

20. Average Ratings of the Interpretability of the three Physiological Responses by Examiners and Raters for Correct and Incorrect Identifications 70

21. Median Number of Uninterpretable Responses in the Records of the Thief, Lookout, and Innocent Suspect for Correctly and Incorrectly Identified Records 72

22. Median Number of Movements Leading to Uninterpretable Responses for Correctly and Incorrectly Identified Subject in the First and Second Sections 74

23. Mean Anxiety Ratings of Subjects While Waiting for the Lie Detector Test 75

24. Mean Anxiety Ratings of Subjects During the Lie Detector Test 75

Table | Page

25. Mean "Guilt Reaction" Ratings of Subjects during the Lie Detector Test 76

26. Types of Criminal Offenses of Subjects at Rehabilitation Clinic 80

27. Distribution of Correct Choices for the Five Raters and the Significance of the Results for both Sessions 83

28. Distribution of Correct Choices for the Four Raters and the Significance of the Results in both Testing Periods . . . 90

29. Accuracy Scores in Detecting Lies (Numbers Test) for the Three Physiological Indices under Four Conditions of Testing. 99

30. Frequencies of Detecting Real Numbers Correctly and Decoy Numbers Incorrectly in the Muscle Tension and Exciting Imagery Conditions 101

31. Accuracy Scores for Detecting the Real, the Decoy, and the Neutral Numbers in the Muscle Tension and Exciting Imagery Conditions 102

ACKNOWLEDGMENTS

The efficient execution of the various tasks which formed the basis for this report is due to the complete cooperation of the men who served as lie detector examiners. These men were: Daniel Baer, John Cross, Everett Delahanty, Rudolph Eckhardt, and John Fangman. Their generous help is sincerely appreciated.

Many burdensome duties were dispatched with unobtrusive smoothness through the skillful supervisory talents of Dr. Baer, who was succeeded in this capacity by Mr. Eckhardt. In the final stages of the report, the assistance of Mr. Donald Sweeney was invaluable.

The aid and encouragement of Dr. Fabian L. Rouke and Dr. Ralph S. Banay is gratefully acknowledged.

At every stage of the planning and execution of these experiments, unstinted support and assistance was graciously given by the following representatives of Rome Air Development Center: Dr. Shelton MacLeod of the Human Engineering Laboratory and Mr. Robert Byrne of the Quick Reaction Capability Laboratory. Their help greatly facilitated the completion of many phases of this research program.

Finally, to all others who have made it possible for this project to attain the generality it possesses, sincere thanks are offered.

SYNOPSIS

The series of experiments which form the basis of this research program on various aspects of lie detection have special reference to feasibility considerations, namely, the possible use of a computer in lie detection.

The experiments involved: a Simulated Theft, Denial-of-Actual-Crime, Denial-of-Classified-Information, and Countermeasure Techniques. Cardiovascular, respiratory and psychogalvanic response systems were monitored, recorded, and evaluated for lie detection capabilities. Questionnaire, Peak-of-Tension, and Association techniques were employed.

Significant results were obtained in the Simulated Theft Experiment. The lie detection examiners were able to differentiate with significant accuracy among the Thief, the Lookout, and the Innocent Suspect. Independent raters, who based their decisions only on the physiological recordings, were able to attain the same degree of accuracy as the examiners. Of the three physiological systems employed, the psychogalvanic response was evaluated with greatest objectivity and was found to yield the most valid results. In the Denial-of-Actual-Crime and in the Denial-of-Classified-Information experiments, success in detecting lies was limited. The Countermeasure Techniques experiment indicated that individuals with a moderate amount of training can contaminate the physiological records with irrelevant responses which enable them to confuse the lie detector operator and thus elude detection. The Countermeasures Techniques experiment points up the possible sources of invalidity in the traditional lie detection procedures.

Despite the positive findings, limitations in objectivity, uncontrolled invalidating influences, and non-standardized instrumentation in this field create a need for further definitive research before the computer can be considered as an integral component of the lie detection decision. However, the need is critical for the objectification of the measurements of various aspects of the physiological patterns used in lie detection interpretations, and computer type programs can be of great service in the solution of this problem.

CHAPTER I

THE LIE DETECTOR: FEASIBILITY FOR COMPUTER ANALYSIS

A. THE COMPUTER PROBLEM

The imaginative and flexible use of the modern computer has given rise to many dramatic advances in technology, in business, and in such practical arts as weather forecasting. The capability of the computer to digest a complex mass of data and to produce, within minutes, a solution to the problem at hand, has stimulated both the interest and ingenuity of many scientists working with intricate and multiply related problems. Certain of these problems, weather forecasting being a case in point, require standby solutions.

It is no surprise that the possible use of a computer in lie detection work has been contemplated in some circles. On two accounts, at least, this would seem to bear further exploration. The first is complexity; the second, speed.

The traditional lie detector, frequently called a polygraph when more than one channel is employed, usually has three recording pens, each monitoring a basic physiological reaction such as respiration or cardiovascular change. These patterns are complex and not too easy to evaluate by ordinary visual analysis, despite the contention of some practicing lie detector operators. The other physiological response, called the psychogalvanic (or electrodermal) response, although simple in form, does not lend itself to the same simplicity in interpretation. With three variables, two of which give rise to highly complex patterns, the problem of integrating these responses so as to yield a single decision is indeed a formidable one for even the most competent and experienced lie detector operator. He makes his measurements by eye, his evaluations by an inner mental weighting process, and his decisions, in large part, by experiential intuitions and hunches. This ia a far cry from the scientific reputation often accorded the lie detection process. Many of the complicated operations which seem to be inherent in the evaluation of polygraph records might better and more easily be performed by a computer.

The ordinary lie detector expert evaluates the charts globally, usually without measuring any responses. Pressure of time is the reason given for not taking precise measurements. It is true that law enforcement agencies like to have answers immediately, and it is also true that the measurement of the physiological responses would take considerable time. This, then, provides a second reason for which a computer might be considered in the processing of lie detection data.

It can be programmed to measure precisely and it can do this quickly.

The computer might be able to perform two additional and related functions for lie detection. First, the outputs of the polygraph could be combined into a multivariate decision function which maximizes accuracy by optimally weighting the different physiological reactions. This would require the precise identification of deception criteria and the development of discriminant weights to be programmed into the computer. In the second place, the computer could be used to discover new criteria of deception that have not as yet been identified in the complicated physiological patterns. Harmonic analysis might reveal aspects of the physiological responses which ordinary human observation could not readily detect. But both of these possible uses assume validity and objectivity in the lie detection procedure, and these characteristics require verification. The feasibility of using the computer in lie detection depends upon the adequacy of the lie detector to help detect lies. What is stressed is the adequcy of the lie detector and not the capability of the operator. For the prime requirement in computerization is that its input be objective and not dependent upon the unique talents of exceptionally gifted operators.

B. THE LIE DETECTION PROBLEM

Lie detectors have been highly publicized in the popular press and sometimes exploited by unscrupulous operators. Nevertheless certain dedicated scientists have reflected their confidence in these procedures by suggesting that they be used in arms control inspection programs (Orear, 1961). Despite the faith many have in lie detectors, there has been very little relevant research on those persons and in those areas where such instruments are most likely to be used (Kubis, 1957).

The fundamental difficulty is lack of standardization. There is disagreement as to the relative values of the physiological responses used in lie detection. There is no uniform set of indices to serve as a criterion for deception. There is an expressed dissatisfaction with the lack of instrumental development (Dana and Barnett, 1957; Dana, 1958). There is a growing awareness of the degree of subjectivity involved in the lie detection reports (Holmes, 1958). This, then, is not the picture of an established science. Rather, it is that of a growing art.

Because of the great increase in the number of men who claim to be lie detector operators, associations have been formed, qualification requirements formulated, and some attempts at professionalization initiated. But despite the acknowledged needs for scientific research

in the field, there is a fairly uniform attitude among the lie detector examiners that their work is an art in which the vital and personal interaction between the operator and the suspect can never be eliminated. It is no surprise, then, to find these men placing great emphasis upon improving interrogation procedures, especially those for eliciting confessions. This may well be a tacit, perhaps subconscious, acknowledgment that lie detector instruments are not as diagnostic as they had expected them to be. However, there has been very little clamor for new instruments and new approaches in a field that has been threatened with stagnation because its growth has not been stimulated by new and imaginative ideas for quite a time.

The major need, then, is a thorough re-examination of the theory, procedures, and techniques utilized in lie detection. The major problem, especially as it concerns the use of the computer, is the objectification of the measurements and the interpretations of lie detector responses. The purpose of this project is to provide some evidence concerning the feasibility of introducing the computer to lie detection work. To attain this objective, major emphasis has been given to evaluation of those basic elements in polygraph examinations which are closely related to objectivity.

C. SPECIFIC AIMS OF THE PRESENT PROJECT

Since computerization involves electrical inputs, the variation in the traditional physiological responses must be evaluated with the use of new transducers. This is especially pertinent to the cardiovascular and respiratory responses which have traditionally relied on the transmission of pressure changes to actuate the recording pens.

The experiments must maintain strict controls on the examiner in order to evaluate the degree of subjectivity in his evaluations. To better achieve this objective several examiners must be used. In this way, the evaluations of a particular examiner could be checked by another who would make his interpretations from the records alone. Finally, the objectivity of the ratings must be determined.

The research plans also must include the use of multiple indicators and multiple approaches. The data obtained from these sources would then be available for a validity study.

Especially important is the problem of contamination in the records, whether it be intentional or unintentional. Some light must be shed on this problem, in the event that deliberately induced distortions can be detected only by the operator.

D. THE PROJECT

To fulfill the aims mentioned above, the project was divided into four phases. The major effort was confined to the Simulated Theft Experiment in which the lie detector examiner and independent raters attempted to differentiate between the Thief, his Lookout, and an Innocent Suspect. Two other pilot experiments were conducted to clarify the problem of realism in lie detection experiments. The objection to laboratory investigations is that the realism of an actual criminal situation can never be fully duplicated. The Denial-of-Actual-Crime Experiment and the Denial-of-Classified-Knowledge Experiment attempt to study the issues involved in the problem of realism.

To study the several elements involved in objectivity, five examiners were trained, for a period of three months, by an experienced lie detector expert. These men were used as operators and as raters, and their scoring, interpretations, and decisions were evaluated for consistency and objectivity.

The polygraph used in the project recorded respiratory, cardiovascular and psychogalvanic response, the traditional physiological measures used in lie detection. These provided multiple indices to be used by the operators and raters in arriving at their decisions. The multiple approach was provided by three methods usually associated with lie detection procedures. The traditional interrogation was done through standard questionnaires. In addition, data were available from a Free Association and a modified Peak of Tension Test.

The final phase involved an experiment to determine the extent to which a suspect could manipulate his reactions and thereby deceive the examiner and elude detection. Several classes of countermeasures were developed and their efficiencies evaluated.

In the following chapters, these phases are given separate treatment. The next chapter, possibly the most important, describes the basic instrumental, examining, and interpretative techniques used in several phases of this project.

CHAPTER II

METHODOLOGY: SIMULATED THEFT EXPERIMENT

A. GENERAL CONSIDERATIONS

The methods and procedures used in the Simulated Theft Experiment possess sufficient generality to be directly applicable to the other experiments described in later chapters. These methods, then, will be thoroughly discussed in this chapter.

The "theft" provides the core experience under examination in the lie detector test. Its description will enable one to follow the temporal course of the experiment, that is, from the commission of the theft to the lie detector examination. In this section the constitutive elements of the test situation--questionnaires, association techniques, and instrumental controls--will be thoroughly examined. Some minor changes in methodology introduced midway through the experiment will be treated in a separate section.

The evaluations of the chart recordings provide the basic data of the experiment. The nature of the criteria of deception and their specific use in the analysis of the records will form a critical section of this chapter.

Of major importance and basic to the validity of the results are the training, ability, and objectivity of the examiners who conducted the lie detector tests. The methods used to estimate the degree of confidence in their ratings and in their decisions will also be described in this section.

The subjects, their behavior and cooperation, and their reactions to the various aspects of the test procedure are of great importance in evaluating the lie detector as an investigative tool. They form one of the three fundamental poles of a highly interacting triad: Examiner-Lie Detector-Suspect.

The concluding sections of this chapter will treat several important problems, such as realism in lie detection experimentation, leakage of information about the details of the experiment, and tension and anxiety in subjects as possible invalidating influences in lie detection tests.

B. THE SIMULATED THEFT SITUATION

Stated simply, the theft situation involved a student (THIEF) who stole the money in a pamphlet rack with the help of another student who served as a LOOKOUT to warn him of approaching students or faculty. Since the "culprits" left the scene of the "crime" by different routes without communicating to each other, the Lookout did not know what and how much was taken from the money chamber of the pamphlet rack. Both Thief and Lookout were later tested by a lie detection EXAMINER. Another student, unknown to the Thief and Lookout, was also tested by the examiner. Since he knew neither who was involved in the crime nor what was stolen, he was an INNOCENT SUSPECT. The order in which these students were examined was randomly determined before the experiment began. At no time after the theft, or before the examination, were the three students of this group in contact or in communication with each other.

The control of the situation was in the hands of a SUPERVISOR who took care of the instructions, maintained the order of testing, and interviewed the students briefly after the lie detector test. The lie detection examiner was given no cues or clues as to the experimental role of the students he examined. His job was to determine this from the records obtained in his lie detector test.

1. Recruitment of Subjects

Students of the college were contacted informally in the cafeteria, at student meetings, or in a class. A brief statement that subjects were needed in a research project, and that they would be paid for their participation in it, usually sufficed to attract their interest. When told that there was an added opportunity to win a substantial money prize, they generally volunteered and an appointment was arranged. If two or more students were contacted at the same time, they were given appointments for different days so that they would not become members of the same group. The majority of students were obtained as subjects on this appointment basis. However, when a subject failed to keep his appointment, it was the duty of the supervisor to recruit a new member to fill out the necessary complement for the group scheduled at that time. Usually the supervisor was helped in this task by an assistant.

Although not all students volunteered for the experiment, no biasing factor could be observed in the groups finally selected. The unpaid volunteer bias was not present. There was no evidence of unusual or morbid interest in psychological experimentation by the stu-

dents who volunteered, nor was there any evidence that fear of psychologists motivated those who did not volunteer. The reasons for declining to be an experimental subject seemed valid and sensible.

The recruitment directions may be found in Appendix A.

2. Role of Supervisor

When the three students began arriving for their appointment, the supervisor assigned each to a separate waiting room. In this way the student who was to be the Innocent Suspect could not know the identity of the other members of the group. One of the group was randomly assigned the role of the Innocent Suspect. The other two students were then given specific instructions about the experiment.

First, the two students drew lots for the task they were to perform, one to be a Thief, the other a Lookout. They were then told that a pamphlet rack was located in the basement and that the Thief was to rob it of its contents. Since the entrance to the men's lavatory was on the same corridor, a Lookout was necessary to warn the Thief of any approaching student or faculty member. They were to develop whatever signals were necessary to safeguard themselves from detection or apprehension by those passing by. In fact, they were warned <u>not</u> to get caught.

So as not to be seen together, they were to return to the testing room by separate routes as soon as the theft had been completed. This instruction also served to prevent communication between the Thief and the Lookout after the crime. In this way the Lookout would not know how much money was in the pamphlet rack nor what else could have been taken. To guarantee this outcome, both the Thief and the Lookout were further instructed not to talk to each other after the crime.

Both subjects were finally told that they would be examined and questioned about the crime. They were instructed to deny any participation in or knowledge of the robbery.

Throughout his instructions the supervisor used such words as "rob", "steal", "danger", and "do not get caught" and never such words as "experiment", "you don't have to worry", and "fake robbery". The purpose was to instill a serious attitude into both subjects and to generate a climate of danger which would be threatening and which they would try to avoid or minimize. That this was actually the case is illustrated by the sudden return of two breathless subjects very soon

after they left to commit the crime. The supervisor was about to assign them for testing when they informed him that they had not yet accomplished their mission and would wait upstairs awhile until they could "shake" the person who was following them. In point of fact, no one was deliberately following them, but they assumed so strong a "robbing" attitude that their feelings of suspicion and guilt endowed an innocent passerby with suspicious and punitive attitudes. Although it was an experimental and not a real situation, the directions, the procedure, and the roles assumed seemed to have induced a personal experience that might very well mirror the attitudes and feelings experienced in a real theft situation. From those who volunteered their feelings, this was an exciting, though not pleasant experience. However, a discussion of the problem of realism in experiments of this type will be reserved for a later section.

Once the "crime" was completed, the supervisor was responsible for the proper routing of the subjects to the examiner in the testing room. At this point the Innocent Suspect had already been instructed as to what was expected of him. He was simply told that a pamphlet rack had been robbed and that he would be questioned about it. But since he had nothing to do with the theft, there was nothing to be concerned about. All he had to do was to answer the questions truthfully.

Each subject was then escorted to the testing room by the supervisor and briefly introduced to the examiner. After the test, the subject returned to the supervisor who interviewed him briefly about his attitudes and reactions to the experiences he had just undergone. The purpose was to obtain additional information about the conditions under which the theft occurred and any unusual circumstances that might have bearing on the validity of the procedure. In addition, the supervisor asked him to fill out a rating scale and then paid him for his participation in the research.

For the actual instructions and directions given by the supervisor, the reader is referred to Appendix A.

3. Role of Examiner

Simply considered, the role of the examiner involved preparing the examination room and the machine, testing the subject, detaching and filing the records obtained, and dismantling the machine.

Five to ten minutes before the subjects arrived for their appointment, the lie detection examiner isolated himself in the examining room

to put the testing equipment in order and to warm up the machine. This also precluded his involvement with the "suspects" before the examination.

Parenthetically, it may be mentioned that the lie detection examiners also served as supervisors for their co-workers. Thus, in the examination of any one group there was always one supervisor and one examiner. When the testing load was particularly heavy, the men inter-changed roles. This team approach provided variety, eliminated fatigue, minimized emergencies, and made it possible to maintain a relatively constant work schedule.

After the brief formal introduction of the subject by the supervisor, the examiner made a brief statement concerning the machine: that it was a lie detector, and that it was efficient in discriminating guilty from innocent individuals. The attachments were placed on the subject--the pneumograph, the finger oncometer, the palmar electrode, and a microphone. After a brief normative period which provided a sample of the subject's reactivity and which also served to test the adequacy of the attachments, the examination began. A brief association test was used first, followed by a questionnaire requiring only YES or NO answers. Next in order was a modified peak of tension test. Finally, a second form of the questionnaire concluded the test. To discourage questioning from the subject after the test, the examiner told him to report immediately to the supervisor who had further instructions for him.

The final task for the examiner was to evaluate the records. This will be discussed fully in the section on the analysis of the records.

For the actual instructions given by the examiner refer to Appendix A.

4. The Obligations of the Subjects

Each subject first filled out a preliminary questionnaire which included many of the background questions used later in the lie detection examination. At this point he signed a statement to the effect that he would not discuss any details of the research project with his family or friends for approximately a year. He was not aware that the lie detection examination would include a question to test whether subjects did discuss any phase of the research with anyone outside the project staff.

After the lie detection test, the Thief and the Lookout filled out a rating scale to assess their tension-anxiety state during the robbery,

while waiting to be tested, and during the actual questioning in the interrogation room. In addition, they were required to estimate how often during the lie detection test they felt they "gave themselves away" by their physiological reactions. Finally, they indicated how many people were in the vicinity while they were "robbing" the pamphlet rack. The Innocent Suspects also filled out a rating scale covering their feelings of anxiety while waiting for the test and during actual interrogation. They were also required to estimate how many of their physiological responses might have been interpreted as guilty reactions.

The preliminary questionnaire and the rating scales may be found in Appendix A.

5. The Theft Situation

The object stolen was money. The reason for this choice was that money is easily handled and pocketed and would be in the Thief's possession while he was being examined. The amount in the coinbox was $1.25, not an unusual amount. The striking feature was in the denomination: a dollar bill and a quarter. The usual coins in a pamphlet rack are nickles and dimes. Since coinboxes are not opened by students, the Thief and the Lookout had to approach the robbery with some degree of uncertainty as to where the coinbox was and how difficult it might be to pry it open. (In the second phase of this experiment, a button was also included in the coinbox.)

The pamphlet rack was chosen because these could be used to display religious pamphlets the sale of which usually provides money for mission work and charity. Stealing from such sources is considered a particularly heinous crime in a student population such as at Fordham.

The location of the rack was deliberately chosen to provide a source of random traffic. In the basement off a stairwell, on the way to the men's lavatory, and near a soda dispensing machine, the pamphlet rack provided a vulnerable situation insofar as detection was concerned. This fear of detection was further emphasized by the directions which stressed "not to get caught."

The theft situation, then, had a risk factor connected with it. There was also uncertainty as to approach, and readily aroused guilt as to the object to be robbed. The amount taken was easily remembered. Since a prize was being offered for those who best fulfilled the requirements of their role, motivation was strong and sustained throughout the course of the experiment. The features of excitement, fear, possible guilt, and reward for success provide many of the psychological elements present in a real crime. The specificity of the criminal act, of

the object taken, and of the roles assumed made it possible to develop uniform interrogation procedures for all subjects. From the design of the experiment, and from all accounts volunteered by the subjects, it is reasonable to assume that the inner experiences of the subjects would not be too different from those experienced in a real theft.

C. THE LIE DETECTION TESTING PROCEDURES

1. The Association Test

The association test consisted of 11 two-syllable words among which were interspersed the critical ones, --BUTTON, LOOKOUT, COIN-BOX. The positions these occupied in the association list were 3, 6 and 9. Thus, each of these words was surrounded by two non-critical ones, the comparison stimuli.

The word BUTTON pertained to the button that was included in the coinbox in addition to the $1.25 in money. Only the Thief would be aware of this. The word LOOKOUT was expected to arouse a pronounced emotional reaction in the subject who was an accomplice to the Thief. The word COINBOX was presumed to have greatest emotional connotation for the Thief. Clearly, the problem of discriminating the Innocent Suspect from the other two posed no difficulty, provided the association procedure was valid. But to discriminate the Thief from the Lookout was expected to be much more difficult because of the identity of their objective (robbery) and the similarity of their feelings (guilt) during the experiment. The word BUTTON was the only one that had the potentiality of clearly evoking an emotional response in the Thief without doing so in the Lookout. To provide an additional source for this discrimination, the modified peak of tension test was added to the procedure, and this is described in the next section.

It was emphasized that all association words contained two syllables. This uniformity of structure was necessary because the critical words had two syllables. Since there might be a possible difference in reactivity to frequently used words in contrast to those with low frequency usage, the "buffers" surrounding the critical words were selected for equivalent word count values in The Teacher's Word Book (Thorndike and Lorge, 1944).

In the test procedure the association list was given twice. It was felt that the validity of the association test would be increased by a second administration.

The test and the associated directions are presented in Appendix A.

2. The Questionnaires

The subjects were alerted to the fact that a number of questions would be asked about the crime that had been committed. They were instructed to answer only by a YES or NO. These questions (and others) were structured into two questionnaires, A and B, which were administered to all subjects. Both questionnaires may be found in Appendix A.

Questionnaire A was composed of 30 questions which were divided into three distinct groups having well defined functions in the examination procedure. One group consisted of the Critical questions. These were concerned with the details of the crime, as for example, "Did you take the money from the coinbox?" There were three critical questions in each questionnaire, and each was used three times, making a total of nine for the critical group. A second group consisted of the Control questions. These were questions of a personal nature, tending to elicit more concern from the person questioned than ordinary, matter-of-fact questions. Considerable research and practical experience has shown that these questions cause greater physiological reactivity than other factual questions. An example of this type was "Is your draft status IA?" A questionnaire contained three questions of this type, each being repeated three times. The function of these questions was to serve as a reference base against which the critical questions could be compared. Thus, in a criminal situation, the use of the question "Did you eat breakfast this morning?" to serve as a control for "Did you murder _____?" would patently elicit significant reactivity to the latter question, even in an innocent suspect. The emotional impacts of eating and murdering are far from equivalent. The third group was composed of Matter-of-Fact questions. These were ordinary factual questions that served to relieve the subject from the emotional pressures that the other questions created. An example of this type was "Do you have any afternoon classes?" As with the other groups, there were three such questions, each used three times in a questionnaire.

In the present experiment, each critical question was preceded by a control and followed by a matter-of-fact question. To prevent anticipation and specific interactions, each control question preceded each critical question only once throughout the questionnaire. This arrangement permitted a non-repetitive triadic pattern of Control-Critical-Matter-of-Fact questions wherein a critical question was always evaluated in terms of its predecessor and successor.

The purpose of the preliminary written questionnaire, given to each subject before the experiment began, was to provide the examiner with a supply of control and matter-of-fact questions, and to familiarize

the subject with the questions he might be asked. He had already been told that he would be questioned about the pamphlet rack. Acquaintance with the questions in the preliminary questionnaire should have tended to equalize the "familiarity" value of all questions finally used in the test procedure.

Questionnaire B was similar in structure to Questionnaire A. Its purpose was to provide additional critical questions describing aspects of the crime other than those treated by Questionnaire A. Since key questions in a crime are frequently repeated or rephrased in subsequent interrogation, the question "Do you have the coinbox money with you?" was used in exactly the same form in Questionnaires A and B.

Questionnaire B differed in one important respect from Questionnaire A. It included a special question which was used at the very end of the test session. This question, "Before today, did anyone tell you the details of this theft?" functioned as a deterrent, in that it served to remind the subject that the test procedure included a check on information leaks. Its use at the end of the test session did not destroy the comparability of the two questionnaires since the special question was not considered an integral part of Questionnaire B.

Occasionally, unusual circumstances make a particular question inappropriate for a subject who cannot give a clear-cut YES or NO response to it. To avoid the tension that would arise from continued repetition of such a question, alternate control and matter-of-fact questions were supplied with each questionnaire. The examiner could use these in place of inappropriate questions. Two additional questions, called energizers, were available in situations where the subject's psychogalvanic reactivity indicated that he was becoming drowsy. In these instances, a question such as "Do you like Bing Crosby as a singer?" was introduced after a particular triad was completed. Such a question, because it involved an evaluative attitude and because it was very different in content from the others, invariably "awakened" the subject and increased his physiological responsivity.

3. The Modified Peak of Tension Test

Questionnaire B was not used immediately after Questionnaire A. A Modified Peak of Tension Test was interpolated between them in order to prevent adaptation effects from affecting the galvanic skin response. The interpolated test tended to preserve alertness and provided the examiner with possible additional evidence as to the guilt-innocence status of the persons tested.

Theoretically, a peak of tension test is one in which the location of the critical stimulus is known or correctly anticipated by the guilty suspect; as a result, his responses tend to increase as the critical stimulus is approached and then sharply decline once that point is passed.

A simple example will serve to illustrate the structure of such a test. Suppose five dollars were stolen. The examiner presents the following stimuli words in order: one dollar, two dollars, three dollars, four dollars, five dollars, six dollars, seven dollars, eight dollars. The suspect will, undoubtedly, say NO to each stimulus, if it is in the form of a question. Or, he may be instructed to give no answer to each of the stimulu. The naive theory associated with this test assumes that the subject will increase in reactivity as the stimuli get closer to the critical issue, namely, the five dollars. At this point, maximum (or peak) responsivity is expected. Thereafter, reactivity should drop in a rather dramatic fashion.

This procedure was modified by having the subject respond with a number just one unit larger than that presented as a stimulus. The numbers used were: 84, 99, 114, 124, 134, 139. The subjects were advised that 99 meant 99 cents etc. Since only the Thief knew the amount of money stolen, he alone was expected to react significantly to 124; all the more so, it was felt, because he had to mention the exact amount taken from the coinbox.

The directions for the test may be found in Appendix A.

D. CHANGES IN PROCEDURE

Although the subjects promised not to talk about the experiment for a period of a year, it was felt that some information leaks might still occur. Consequently, some minor changes were made in the procedure at the half way point in the experiment. One purpose was to disturb the confidence a subject might have in any "illegal" information he received. At the same time, the changes were to incorporate whatever improvements could be made in the test procedure. However, they were not to be so drastic that comparability between the first and the second amended phase would be meaningless.

The original plans called for 100 experimental groups. Because of the relative complexity of the experiment and the emotional tension it engendered, it was felt that many groups would be lost for failure to follow directions. Anticipated breakdowns in instrumentation were expected to bring about the elimination of other groups. For these reasons, 125 groups were scheduled, a number sufficiently large to guarantee a final total of 100 experimental units. Changes in

procedure were introduced after the sixtieth group was tested, approximately midway in the experiment.

1. Changes in Questionnaires

To better appreciate the changes in the questionnaire, the critical questions are reproduced below for both sections of the experiment.

Critical Questions in First Section

Questionnaire A

1. Do you know who took the money from the coinbox?
2. Did you take the money from the coinbox?
3. Do you have the coinbox money on your person?

Questionnaire B

1. Are you completely innocent of this crime?
2. Did you act as a lookout for the thief?
3. Do you have the coinbox money on your person?

Critical Questions in Second Section

Questionnaire A

1. Were you an accomplice to the thief?
2. Did you take the money from the coinbox?
3. Do you have the coinbox money with you?

Questionnaire B

1. Do you know how much money was in the coinbox?
2. Did you act as a lookout for the thief?
3. Do you have the coinbox money with you?

The elimination of the first critical question in Questionnaire B (First Section) was the major change. It was the only critical question that had to be answered by a YES. Its exclusion from the second section of the experiment made it possible for NO to be the uniform response to all critical questions relating to the theft. In Questionnaire A the change in the first question served to provide a logically clearer discrimination between the Thief and the Lookout. The change in the third question was stylistic in character. No changes were made in the control or matter-of-fact questions.

2. Change in Coinbox Contents

To provide the robbery with an element of which only the Thief would have knowledge, a button larger than a quarter was introduced into the coinbox along with the $1.25. The size and color of the button were readily distinguishable from a silver coin. It was anticipated that this would provide uniqueness to the Thief's experience. It would also provide the Free Association Test with a better stimulus for discriminating between the Thief and the Lookout.

3. Changes in the Free Association Test

In the first half of the experiment, the critical words in the Free Association Test were: pamphlet, lookout, coinbox. With the change in the coinbox contents, the word button was substituted for pamphlet. A further change in procedure was to administer the test twice, one administration immediately following the other. Under these conditions, the second application could also serve in a peak of tension capacity.

There is no doubt that the changes introduced into the second phase of the experiment provided better controls and greater opportunities for correct discrimination among the three types of suspects. But basically the experiment remained the same. For this reason, the descriptions of the techniques referred mainly to the form in which they were used in the second section.

E. INSTRUMENTATION

The three standard physiological responses used in lie detection are blood pressure and pulse indications, respiration, and the psychogalvanic response. When packaged as one unit, the three instruments measuring these reactions are called a Polygraph. Several different companies manufacture polygraphs which record all three responses on one chart. The pens that record blood pressure and respiration are in mechanical contact with the pressure chambers that are connected with an arm cuff and with a rubber pneumograph.

Since one of the objectives of this research was to determine the feasibility and advisability of computer coupling with the traditional lie detector, such reactions had to be transformed to voltage changes. The psychogalvanic response is usually recorded in the desired form, but transducers had to be used for the pressure changes generated by the cardiovascular system and by the lungs. A further change, one that has been long desired by most lie detection operators, was introduced. The

usual medical pressure cuff is very uncomfortable and cannot be used in prolonged interrogation because of possible damage to the blood vessels in the arm. For this reason, it was decided to use a volume plethysmograph which would yield recordings from a single finger without discomfort to the subject. There is some indication that the plethysmograph might serve as a reasonably adequate substitute for the blood pressure cuff (Dana, 1958; Davis, 1961).

The Model 5 Polygraph, manufactured by the Grass Instrument Company, was used in this research. A photograph of the Grass Polygraph is included in Appendix F.

1. The Psychogalvanic Unit

A low level, high gain, D. C. pre-amplifier (5P1) recorded the psychogalvanic response. A 50 microamp current was passed through the electrodes with impedance of 2.3 megohms. Under conditions of maximum sensitivity, a pen deflection of one centimeter was equivalent to a 100 ohm change. High stability was indicated by a maximum drift of 3 microvolts per hour. Balance voltages had an accuracy of 1 per cent.

Two nickel silver electrodes (2 1/4" x 1 1/2") were attached to the palmar surface of each hand. Rubber straps were used to keep the electrodes attached to the palms. The subject rested each hand comfortably upon the wide flat arms of a special chair used for interrogation. Since the subject was warned against arm or finger movements and since the position of his arms was quite comfortable, relatively few movements were made and no records were rejected because of such artifacts.

2. The Plethysmograph

To measure and record volume changes in peripheral circulation, a metal finger cuff, called an oncometer, was slipped over the middle finger of the left hand. The changes in volume were measured by a pressure transducer (Grass PT 5) which featured rugged performance but high sensitivity at low pressures. It was developed primarily for digital plethysmography. A low gain, high input impedance, capacity coupled amplifier (Grass 5P2) was used with the pressure transducer. Its maximum sensitivity was 100 millivolts per centimeter. No significant noise voltage was developed at the amplification used. When excessive movements blocked the amplifier, a trace restorer instantly returned the pen to its normal position. This amplifier was designed for use with ceramic transducers to record changes in pressure and movement.

3. Respiratory Changes

A Phipps & Bird Bellows pneumograph was used to record respiratory changes. With it, a ceramic volume change transducer (Grass CP1) was connected to the input of a high-impedance amplifier (Grass 5P2) to measure variations in breathing. Since absolute levels were not required in recording these changes, use of this transducer was appropriate and convenient, eliminating input balancing operations.

4. Attaching the Subject to the Instrument

When the subject was comfortably seated in the large arm chair, the pneumograph was the first to be attached, followed by the electrodes and the oncometer. When the amplifying and recording systems were in proper working order and when a sample of the subject's typical reactivity was obtained, he was alerted for the start of the test.

The subject was seated facing away from the examiner so that neither the examiner nor the subject would be affected by each other's facial expressions. A photograph of a person sitting in the chair and attached to the instrument is contained in Appendix F.

5. The Examining Room

The lie detection examination took place in a relatively small interviewing room which was fitted with a one way observation mirror. All operations and activity in the room could be observed readily and without the knowledge of the subject. With a two way communications system between the examining and observing rooms, the interview between the examiner and the suspect was available for tape recording.

Two interviewing rooms were used by the examiners. This arrangement created no difficulties, since the polygraph and tape recorders were portable.

F. ANALYZING AND SCORING THE RECORDS

1. The Problem of the Diagnostic Indicator

In the early history of lie detection, many investigators were of the impression that the physiological counterpart of the conscious lie would be specific and distinctive. The search at that time was for a single criterion such as shallowness in breathing, rise in systolic blood pressure, and increase in psychogalvanic response. This criterion, whatever form it took, was assumed to be a diagnostic sign manifested

by all who lied. In time, the hypothesis of specificity (of criterion) and universality (of application) was abandoned. It was realized that different people gave varied physiological reactions in lying situations. In fact, one investigator writes the following:

> "In the cardio-sphygmograph tracing the following basic changes may be indicative of deception:
>
> a. Increase in blood pressure.
> b. Decrease in blood pressure.
> c. Increase in pulse rate.
> d. Decrease in pulse rate.
> e. Increase in amplitude.
> f. Decrease in amplitude. . .
>
> In the galvanograph tracing, the following basic changes may be indicative of deception:
>
> a. Vertical rise at point of deception. . .
> b. Drop at point of deception (in non-centering galvanometers)." (Joseph, 1957, pp. 94-95)

The implications of such a listing may prove rather disturbing to some scientists. As one reader facetiously remarked "It makes no difference how the suspect reacts. He will be found guilty." But the clue to the proper interpretation of Joseph's extensive listing is the presence of the category "general pattern changes." It is listed as an index of deception within each of the three basic physiological reactions. This means that a particular individual may exhibit a unique pattern change when he lies, and this change may be very different from that found in another person. Each person's lie, then, has a specific signature. If an operator could obtain a sample or two of the physiological patterns of a suspect's lying response before the actual test is administered, he would be in a better position to evaluate the records of the suspect while he is questioned about the crime under investigation.

However, lie detection examiners do not have such samples of genuine lying behavior prior to the test they may have to administer to the suspect. So-called control tests do not provide adequate criteria nor sufficient control. As a result, the ordinary lie detector operator is forced to use the criterion of "general pattern changes" to evaluate the responses at the point where the critical question is asked. It is the criterion which has been used in the present study.

2. The Rating of the Critical Responses

The strengths (and weaknesses) of lie detection depend more or less exclusively on the emotional resonance the suspect exhibits to questions (or stimuli) relating to an actual crime. This theoretical structure assumes that the guilty individual manifests a pronounced physiological disturbance in denying his guilt, while the innocent suspect shows no such disturbance in asserting his innocence. However, gradations exist in these disturbances, thus making it possible to develop a rating scale with more than two categories.

There are, to be sure, at least two objective measurements that can be made on each physiological response mode. Thus, for cardiovascular changes one may measure the rise in the pressure curve and in the pulse rate. Respiratory responses differ in amplitude and frequency. As for psychogalvanic responses, amplitude and return time provide precise indices for measurement. But the experienced examiner, in his interpretation of the "general pattern change" that may occur in any of the three pasic physiological systems, notices many more variables. For this reason it was decided to have examiner and rater evaluate each critical response in global terms. Since responses varied in their diagnostic significance, a rating scale was used with provisions for four different categories. Thus, each critical response, when compared with its predecessor and with its successor, was rated as either:

(a) very significant
(b) significant
(c) doubtfully significant
(d) non-significant.

If, for some reason, the critical response could not be compared validly with its neighbors (e.g. because of movement), a "non-interpretable" category was used.

The critical response was called "non-significant" if its pattern structure was no different from that of the responses surrounding it. Thus, in the case of the psychogalvanic response, if the height of the curve for the critical question was not appreciably larger than the others in the triad, it was categorized as "non-significant." The designation "significant" was applied when the critical response was clearly different from the others, and when the rater was willing to consider it as an indicator of lying. Again, to use the psychogalvanic reaction as an illustration, a "significant" rating would be assigned to the critical response if its magnitude was clearly greater than that for the surrounding responses. The "very significant" category was reserved for those

physiological disturbances that were unusually large (again, in comparison to those in the triad) and very outstanding. These would comprise the "textbook" illustrations of what a real lie should look like in its physiological representations.

The "doubtfully significant" category was introduced to aid the rater in his discriminations and to stress the continuity between "non-significant" and "significant" responses. In the context of the interpretative system described above, the "doubtfully significant" response to the critical question was greater than the surrounding responses, yet not sufficiently large to enable the rater to assert with confidence that the suspect was lying.

For purposes of illustration, changes in the magnitude of the physiological pattern were emphasized. Such differences in size were readily observable in the psychogalvanic reaction which was used as an example. The respiratory and plethysmographic patterns, however, were much more complex. A change in any aspect of the physiological pattern ("general pattern change") was considered in the determination of the "significance" or "non-significance" of the response for lie detection purposes. The criteria for such determinations were developed in the training sessions. They remained relatively uniform for all examiners throughout the experiment.

Each examiner assigned ratings to every critical response in the records. For any one critical question there were three ratings, one each for the three physiological reactions. To facilitate the rating and recording, a special chart was constructed. A copy of this rating chart is included in Appendix E.

3. The Scoring System

For numerical analysis, the ratings were transformed into quantitative terms. The scoring system adopted was unusually simple and reflected the varying degrees of importance the raters assigned to the verbal categories they had been using. The numbers 3, 2, 1 and O were assigned to the "very significant", "significant", "doubtfully significant", and "non-significant" categories, respectively. The scaling was ordinal, at the least. Some raters felt that this assignment of numerical values to the ratings may have achieved interval scaling--insofar as their own ratings were concerned. While this is a dubious assumption for the complicated patterns of respiratory and blood volume changes, it is not unreasonable for the psychogalvanic responses, the magnitudes of which were scaled by the recording chart. Since the raters felt that the scores were reasonable estimates of the discriminating values they themselves associated with their ratings, and since the training of the raters was both intensive and uniform, thus assuring equivalent criteria of rating, it

was deemed appropriate to use these scores as if they were measures on an interval scale. These numbers, then, could be added together for the same rater and for the different raters in the several analyses used in this study.

4. The Decision: Reliability-Objectivity Considerations

In actual lie detection practice the determination of guilt is never made on the basis of one response alone. The decision that a suspect is lying is made on the basis of many reactions to many critical questions (or stimuli). Theoretically, the liar is expected to give many "significant" responses; the truthful person, none or few. Consequently, a "significant" response does not necessarily indicate a lie, but it has a definite probability of being associated with lying. However, when a "significant" reaction is consistently associated with the same critical question, it is usually presumed that the person is lying to that question. Such a conclusion is particularly appropriate if the suspect shows no similar emotional involvement to the control questions.

The above considerations stress matters of reliability and consistency. In this experiment there was a similar emphasis, and many stimuli and several different approaches were used to provide a basis for a valid and reliable decision. Thus, there were two questionnaires, a free association test, and a peak of tension test. The questionnaires included many different critical questions in order to encompass as many facets of the "crime" as was feasible. To study the consistency of the "guilt reaction", each critical question was repeated three times within the same questionnaire.

In the evaluation process, the records of each experimental group were analyzed and rated as a unit. In other words, the examiner studied at one time the records of the three participants in the simulated theft: Thief, Lookout, and Innocent Suspect. After the records were rated, the examiner (or rater) evaluated the ratings and made his decision as to the role the subject assumed in the experiment. Inevitably, this was a comparative process in that the decision was made only after the records of all three members of an experimental group were rated.

In addition to the examiner, two independent raters were assigned to the same set of records. Their ratings and evaluations were made without any knowledge of the role the subject had in the experiment. The examiner's ratings were "blind" too, in the sense that he did not know precisely which of his subjects was the Thief, which the Lookout, and which the Innocent Suspect. However, he could definitely associate the chart with the person who produced it. Consequently, any conscious or

unconscious biases, which may have been formed in observing and testing the subject, may have influenced the examiner's ratings and evaluations. But the two independent raters were completely free from such biases. They did not know or see the subjects who were examined; further, they had only the objective records to rate and evaluate.

The use of two independent raters was particulary pertinent in evaluating the objectivity of the rating procedure. In addition, some estimate of possible examiner bias could be obtained by comparing the agreement between the two independent raters and that between the examiner and independent rater for the same set of records. If the personal contact between examiner and suspect was an important influence upon the examiner's ratings, one would expect his evaluation to agree less frequently with a rater than the evaluations between the two independent raters.

G. DISCRIMINANT ANALYSIS

In evaluating the polygraph records, the examiner (or rater) had to differentiate among the roles assumed by three subjects in each experimental group. To identify the Thief, the examiner had to differentiate his responses from those of the Lookout and Innocent Suspect. Similarly, the identification of the Lookout was possible because his responses could be discriminated from those of the other two members of the group. The final decision, involving a discrimination among three subjects, was undoubtedly based on such subsidiary comparisons as those between

a. the Thief and the Innocent Suspect
b. the Lookout and the Innocent Suspect
c. The Thief and the Lookout

It is clear that once two members of an experimental group were identified by the examiner, the third was necessarily determined, --even without an analysis of his records. Though this is a theoretical possibility, it was not the procedure followed. All three records were analyzed, evaluated, and compared with one another before the examiner made his final decision.

To make his discriminations among the subjects, the examiner had to evaluate each of the three physiological tracings in each record. Since these physiological systems are not perfectly correlated and since each is assumed to contribute a different aspect of the physiological disturbance involved in lying, the major problem becomes a matter of how much importance is to be attached to each physiological tracing. In other words, what are the relative weights to be

given to the pneumograph, the plethysmograph, and the psychogalvanometer when these are used as components in a lie detecting instrument? The examiner, no doubt, makes such differential weightings, consciously or unconsciously, when he analyses the polygraph chart. Such personal weighting systems undoubtedly change from examiner to examiner and from one time to another for the same examiner. There is need, then, for an objective weighting system that would assign appropriate numerical values to each of the physiological tracings in order to maximize the accuracy with which discriminations can be made.

The purpose of this section is to describe the construction of discriminants, or mathematical functions, which optimally classify objects into the groups where they most likely belong. There are simple and multiple discriminants. These, in turn, may be linear or non-linear. Because of special interest in the dichotomous classification, the discussion will be concerned with simple linear discriminants. These will provide an optimal weighting of the three physiological indices in order to differentiate most accurately between two classes of subjects, as, for example, between the Thief and the Innocent Suspect.

The primary requirement is a scoring system which would most likely yield different values for the Thief, the Lookout and the Innocent Suspect. If, for example, such scores are developed and they are consistently and uniformly different for these three groups, they can be used in an objective discriminant function. Under these conditions, the mathematical function could be used to make the decision, that is, it would objectively identify the subject's role in the experiment.

1. Thief-Innocent Discriminant Score

The first step was to develop a rational scoring scheme for those combinations of questions that would most likely maximize the accuracy of thief and innocent discriminations. To illustrate the procedure, the questionnaires ir the second phase of the experiment are used. The critical questions from both questionnaires are listed and labelled for easy reference in the discussion that follows.

Questionnaire A

I. Were you an accomplice to the thief?
II. Did you take the money from the coinbox?
III. Do you have the coinbox money with you?

Questionnaire B

I. Do you know how much money was in the coinbox?
II. Did you act as lookout for the thief?
III. Do you have the coinbox money with you?

It is fairly obvious that questions AI and BII are not useful for a Thief-Innocent discriminant because both the Thief and Innocent Suspect answer these questions truthfully when they deny being an accomplice to the Thief. The remaining four questions clearly differentiate the Thief from the Innocent Suspect. In answering NO to these questions, the Thief is lying and the Innocent Suspect is telling the truth. These four questions, then, form the basis for the Thief-Innocent discriminant. In terms of the scoring system (3, 2, 1 and 0), maximum discrimination would be attained if one person got a score of +36 (4 questions, 3 replications of each, and a maximum score of +3 for each response), while another got a score of zero (0 for each response). Such maximum differentiation is possible for each physiological response mode. Consequently, each person would obtain three Thief-Innocent scores, one for the pneumograph, one for the plethysmograph, and one for the psychogalvanometer. Over and beyond the value of such scores for the discriminant function, they provide an opportunity of evaluating the discriminating power of each of the physiological measures.

2. Lookout-Innocent Discriminant Score

The development of the Lookout-Innocent discriminant score was based on the same principles. Here, only two of the questions could be used, namely, AI and BII, for only on these two questions was the Lookout lying while the Innocent Suspect was telling the truth. For this discrimination, the maximum value is +18 (2 questions, 3 replications each, and +3 for each response); at the other extreme, the minimum score is zero (0 for each response to these two questions). Thus, a high score would be interpreted as identifying the Lookout while a low score would tend to be associated with the Innocent Suspect. Again, there were three Lookout-Innocent scores for each person, one for each of the three physiological measures.

3. Thief-Lookout Discriminant Score

In the Thief-Lookout score, positive and negative values were used in the rational weighting procedure in order to maximize the response differences between the Thief and Lookout. The introduction of negative weights was necessary if all questions were to be used. And

since it was anticipated that the discrimination between the Thief and Lookout would be most difficult to make, as many questions as possible were utilized in the construction of this scoring system. For it is rather obvious that the Lookout might consider himself as much a Thief as the person who took the money from the coinbox. In the eyes of the law, they would be equally guilty if this were a real crime. Consequently, the guilt reaction in the Lookout was expected to overflow to the "thief questions"; and similarly, some guilt reactions might be expected from the Thief when questioned about being an accomplice.

If one examines the critical questions for their capability to discriminate between the Thief and the Lookout, it is rather obvious that the Thief would be identified with greatest accuracy if he reacted very significantly to AII, AIII, BI, BIII, but non-significantly to AI and BII. Conversely, the Lookout would be picked out rather easily if he gave very significant reactions to AI and BII but non-significant reactions to AII, AIII, BI, and BIII. To increase the differentiating power of the questions, special treatment was given to AI and BII. Thus, if a non-significant reaction was given to AI, the response was scored +I. But if any of the "high" reactions (very significant, significant, doubtfully significant) were given to this question, they were scored -1. Similarly, question BII received a Thief-Lookout score of +1 if the rating was non-significant; but if the rating was higher than the non-significant category, a score of -I was given to that rating. The purpose of the negative weighting procedure was to distinguish between a Thief who reacted low (or non-significantly) to a question like "Did you act as a lookout?" from the Thief who reacted highly (i.e., with some degree of significance) to the same question. The remaining questions (AII, AIII, BI, BIII) were scored in the usual fashion, with values of +3, +2, +1, 0 corresponding to ratings of "very significant", "significant", "doubtfully significant", and "non-significant". The Thief-Lookout scores range from a maximum of +42 to a minimum of -6, with the high positive scores identifying the Thief and the low scores pointing to the Lookout.

4. The Discriminant Function

From the above analysis there emerge three different sets of scores. Each individual in a group has a Thief-Innocent score for each of the physiological indices. From an a priori point of view, it should be easiest to discriminate between the responses of thieves and innocent subjects; and the accuracy of these scores should be highest. The least accurate discrimination should come from the application of the Thief-Lookout scores for the reasons mentioned earlier. The Lookout-Innocent scores are based on only two questions and the range of these scores is no more than half that for the other two. Despite

the limitation in range, it was anticipated that these scores would be more discriminating than the Thief-Lookout scores.

Once the scoring was completed, the mathematical analysis involved in computing a discriminant function was simple and straightforward. The process first involved classifying each individual according to the actual role he assumed in the experiment. Then, the two roles for which a discriminant function was to be developed were isolated for analysis. The discriminant function was computed for such dichotomous data, yielding the appropriate multipliers (weights) for the respiration, finger volume and psychogalvanic scores. These weights were then used on the data from which they were obtained in order to ascertain the relative success attained by the mathematical function in classifying the individuals correctly.

The general scientific goal was to develop predicting or classifying functions that remain invariant under a wide set of conditions. Thus, it was important to see how variable such discriminant functions were from rater to rater. To realize this objective, three discriminants were developed from the same set of data, one for the examiner and one for each rater. The variations in the discriminant functions for each such triplet of raters were studied. Unfortunately, the sample sizes were small, and it was anticipated that the fluctuations in the discriminants would be large and would mask trends that might emerge in larger samples.

The second kind of invariance that should be tested is the accuracy of the discriminant when applied to a new sample. This is essentially the notion of cross-validation. It was not feasible in this study because the samples for a single rater were too small, and because the data were not objectively measured. Rated data were used and differences among raters preclude the valid use of a discriminant derived from one set of data and applied to that belonging to another rater. The greater the degree of rater objectivity, however, the greater the feasibility of such crossover discriminants.

Despite the difficulties enumerated above, both the discriminant scores and the discriminant functions were used to differentiate between the Thief and Innocent Suspect, the Lookout and Innocent Suspect, and the Thief and Lookout. The accuracy of these objective decision procedures was compared with the accuracy attained by the examiners and raters. In addition, the lie detection capability of each physiological index was evaluated in terms of the accuracy with which it classified the subjects according to their correct roles.

H. SUBJECTS

The subjects of this experiment were male college students. The average age of the sample was 20 years; the standard deviation was 1.6 years. The distribution of these subjects according to class year is presented in Table 1. There was no apparent selection factor operating

TABLE 1
DISTRIBUTION OF SUBJECTS
BY CLASS YEAR

CLASS	FREQUENCY
Senior	77
Junior	81
Sophomore	95
Freshman	83
TOTAL	336

in the recruitment of subjects. The distribution of subjects indicates a reasonably good representation of the four classes.

I. THE EXAMINERS

The five examiners were all psychologists. Three were instructors in psychology and the others were graduate students in a doctorate program. They ranged from 23 to 30 years of age. All were mature, capable men who undertook their work as examiners with serious intent and careful preparation.

To guarantee uniformity of examining technique and consistency in evaluating records, the examiners were given a three months training course in the theory and practice of lie detection. This consisted in selected readings, demonstrations, and practice. They were required to prepare questionnaires, develop experimental test procedures, and analyze lie detection charts obtained in actual criminal situations. The course of instruction was conducted by a professional psychologist who has been an active lie detection operator for approximately 20 years.

All types of lie detection procedures were demonstrated, with the final month of training assigned to the Grass Polygraph which they were to use in the experiment. All examiners were checked out for efficiency by the present investigator in a series of tests closely approximating the conditions of the Simulated Theft Experiment.

J. CONTROLS ON EXAMINERS AND SUBJECTS

1. Examiners

To prevent hasty and unreliable evaluations, examiners were supplied with a rating form which directed them to evaluate every critical question and association stimulus in each record. After this rating, the examiners (and raters) made their decisions as to the thief-lookout-innocence status of the subjects. Subsequent to this, they filled out a final rating form to express: (a) the confidence they had in their decision; (b) the ease with which the records could be interpreted; (c) the help they felt they received from the association and modified peak-of-tension tests; and (d) the influence the behavior of the subject had on their decisions (examiner only). This procedure prevented hurried, ill-considered evaluations and provided some additional information from the examiner on several aspects of his work. In particular, the relation of confidence to the correctness (or incorrectness) of the decision would seem to be a critical point both in the training of examiners and in the value of their testimony in courts of law.

Copies of the several rating forms used by the examiners may be found in Appendix E.

2. Subjects

In addition to signing a pledge not to discuss the experiment with other people, the subject was later questioned by the examiner as to whether anyone gave him specific information about the details of the test procedure. Since this questioning was done while the subject was attached to the polygraph, it was felt that this experience would reinforce the seriousness with which the project was to be considered and would also discourage any tendencies on the subject's part to talk about the experiment. The recruiting directions, too, stressed the requirement of not talking about the experimental procedure and offered the subject an opportunity of winning a substantial prize for following directions faithfully. In this way, it was hoped to control the problem of "information leaks." From all available evidence, during and after the experiment, it was possible to discount any serious leakage problems.

There still remained the question of how involved the subject would be in the experiment as a whole and in the role he played. To obtain an estimate of the attitudes and feelings the subjects experienced during the experiment, the participants in the theft were required to fill out a rating scale immediately after the lie detection examination. They were asked to indicate how tense and excited they were while waiting their turn to be examined and while being questioned by the examiner about the theft. They also estimated the degree to which they may have given themselves away when they denied their participation in the theft. As for their experience during the theft itself, they rated their anxiety and tension states while robbing the pamphlet rack and reported on the number of people who passed by during that critical interval.

The innocent suspects filled out a similar questionnaire. The items pertaining to the act of robbery were, of course, eliminated for these subjects.

These rating procedures were introduced not only for an estimate of the subject's involvement in the experiment. They were also an approach towards understanding the mental state of a person, when involved in an anxiety and guilt producing activity which he might later deny under interrogation. The various rating scales used for the three types of subjects are reproduced in Appendix A.

K. THE PROBLEM OF REALISM

The emphasis in a feasibility study is application and this, in turn, implies careful scrutiny of the problem of realism. The more realistic an experiment, the greater the confidence in the applicability of the results to comparable situations. New problems, however, emerge in the simulation of criminal activity devised for lie detection experiments. In these situations, morality and professional ethics are among the most important issues an experimenter has to face. One cannot have subjects commit a crime in order to evaluate the behavior of criminals in a scientific manner. One cannot force a subject to do wrong just to study one or more facets of human behavior experimentally.

In this particular experiment, the role each subject assumed was associated with a number of realistic features that were designed to reproduce the mental excitement, anxiety, and guilt expected of a thief in a real life situation. The directions and experimental plan were designed to make the subjects behave like thieves on the mission to which they were assigned. It was felt that the procedure enabled the subjects to live through an inner emotional experience closely approximating that which an actual thief would undergo.

That the Simulated Theft Experiment approximated realistic conditions has been briefly discussed in an earlier section of this chapter. Additional evidence comes from the reactions of a number of subjects in the various stages of the experiment. When acquainted with the type of tasks involved, two subjects declined to continue with the experiment because they would not have felt right in following out instructions. Another subject, the Thief in the experiment, fainted at the beginning of the test situation because of the excitement generated by his "criminal" activity and by the tension evoked by the examination. Still another subject, a Lookout, unconsciously (or consciously) divulged the fact that he was involved in the theft. A large proportion felt relieved in getting rid of the "stolen money" after the completion of the interrogation. Two Thieves, in their excitement, did not remember how much money they had taken. These and other reported attitudes indicated that the experiment generated excitement, anxiety, and guilt. The directions and situation, too, were designed to facilitate personal involvement by the subjects, who, in the past pace of the experiment, did not have much of an opportunity to consider the non-real (or artificial) aspects of the theft situation.

Despite the favorable realistic features of the experiment, new situations were planned which would approximate other aspects of a criminal's behavior when questioned with a lie detector. The strong motivation in the criminal to withhold information was, in all probability, not the same type of motivation the experimental subject had in denying his participation in the simulated crime. Consequently, a special rehabilitation group was tested on matters pertaining to previous criminal behavior which they would ordinarily not discuss with unauthorized personnel. It was felt that this group would be sufficiently motivated to deny former criminal activity, especially since they were in the process of rebuilding their lives along legitimate, normal paths. An additional group was included to represent those who, by selection and training, would be sufficiently motivated not to reveal classified information. Large numbers of persons are engaged in work of a classified nature and there are relatively many people having access to classified information. A sample from among these was used in the present research.

Both samples were highly motivated to withhold information for various reasons, either personal or related to national security. Both samples were included because they seemed to afford a test of the lie detection procedures in a more realistic situation. There remained one additional facet to this problem. While motivation, especially that which aims to maintain personal security, is important in experiments with the lie detector, the problem of the defense maneuvers such

a highly motivated person might take needs special attention and study. In other words, the criminal under interrogation might be so highly motivated to withhold incriminating information, that he would take all possible measures to evade detection or to nullify the value of the lie detector records. In other words, he would try to "fool the machine" in every way possible. This is the problem of countermeasures which is explored on a pilot basis in the present research. Since there are special methodological problems involved in experimenting with these three groups, a separate chapter will be devoted to each of them. They are mentioned in this section in order to complete the discussion of the ways in which the problem of realism was handled in the present research.

L. INCOMPLETE AND DISCARDED RECORDS

In any experiment where directions are either lengthy or involved and where considerable freedom is given to subjects to execute a task, some failures may be expected. These may be due to slips in supervision or to misinterpretation of directions. Instrument failure is another source for rejecting records. Finally, some characteristics of the subject may violate the homogeneity requirements of the sample. Table 2 presents the number of experiments which were incomplete or discarded and the reasons for such action.

TABLE 2
CLASSIFICATION OF INCOMPLETE AND DISCARDED RECORDS

FAILURE	NUMBER
Supervision	3
Instrumentation	2
Subject	2
Instruction	5
TOTAL	12

Among the supervision failures was one in which the money was not placed in the pamphlet rack and consequently no money was stolen. The other two cases were clerical failures. The order in which the subjects were tested was not recorded, and there was no way in which the

appropriate name could be associated with its record.

The instrumental failures were trivial in nature. A break in the electrode connection prevented the completion of one experiment. In the other, the oncometers could not be fitted to one subject's fingers.

Of the two subject failures, one experiment was discarded because a subject had been tested extensively on the lie detector in an actual criminal case. In the other case, the subject fainted while attached to the polygraph. He was not permitted to continue the test but was sent to the infirmary.

Instruction failures were of several types. Two subjects gave themselves away during the test, effectively revealing their roles to the examiners. In another group, the subjects misinterpreted the directions to the extent that they felt that the Thief and Lookout should falsely admit to the other's role. Two other subjects attempted to manipulate the physiological controls by movements and changes in the breathing rate.

CHAPTER III

RESULTS: SIMULATED THEFT EXPERIMENT

The most extensive phase of this research was the Simulated Theft experiment. In many ways it was the best controlled of the three experiments on different subject groups. The Denial-of-Actual-Crime and the Denial-of-Classified Knowledge experiments were conducted away from the Fordham Laboratories. Consequently, space arrangements and examining conditions could not be the same as at Fordham, although they were adequate for the purposes of the experiment. A more important consideration concerns the statistical reliability of the results. Since the number of subjects in the Simulated Theft experiments far exceeded that in the others, the results are more stable and engender greater confidence in the suggested conclusions. For these reasons, the treatment of results in this chapter will be more extensive.

The first part of this chapter will deal with a general overview of the results, both in terms of accuracy of detection and in terms of inter-rater consistency. Then the accuracy of the five examiners will be compared with that they attained when serving as raters. Since the records were evaluated by raters without any knowledge of who the subjects were or what roles they played, the data from such "blind" analyses will provide an estimate of accuracy uninfluenced by any subjective impressions of "deceptive" behavior the examiner may have observed in testing the subjects. The agreement between examiners and raters and that between the raters themselves should provide another index of the relative objectivity of the lie detection results. Along similar lines, discriminant analyses will provide still other estimates of the accuracy of the lie detection procedure to differentiate the Thief from the Innocent Suspect, the Thief from the Lookout, and the Lookout from the Innocent Suspect. The relative accuracy of the three physiological indices (respiration, blood volume change, and electro-dermal response) will be discussed at this point. The several auxiliary topics concluding the chapter will include an analysis of the errors made by the examiners and raters; an evaluation of the anxiety and tension experienced by the subjects during the experiment; and a study of the confidence of examiners and raters in their decisions.

A. GENERAL OVERVIEW

There were two sections of the Simulated Theft experiment. In the first section, 54 groups, each consisting of three subjects, were tested. A slight change in the experimental procedure was made in the second section which involved 58 comparable groups. To evaluate possible trends during the progress of the experiment, the data were divided into succes-

sive thirds. This fractionation was done for both sections of the Simulated Theft experiment. In order to have equivalent comparisons between both sections, the thirds comprised 18 groups. The last 4 groups of the second section were kept separate in this trend analysis.

There were three members in each group tested by the examiner: the Thief, the Lookout, and the Innocent Suspect. Not knowing the role each subject played, the examiner, on the basis of pure chance, could use six different sets of decisions, namely, ILT, ITL, LIT, LTI, TLI, TIL. ILT symbolized the decision that the first subject was the Innocent Suspect, the second the Lookout, and the third the Thief. Thus, the examiner's accuracy of detection can be evaluated against the chance expectation of 1/6. In this general overview, the examiner's accuracy score will be the number of groups judged with complete accuracy.

Similarly, in discussing the rater's accuracy, the same criterion will be applied, namely, the number of groups in which the role of each subject was correctly identified.

1. Accuracy of Examiners and Raters

Every record was evaluated by the examiner and independently by two raters. The examiner's evaluation was, naturally enough, not independent of his observation of the subject's behavior during the testing period. Only the records themselves were available to the raters. If the examiners were systematically influenced either positively or negatively by the behavior of the subjects, a difference in accuracy between examiners and raters should be observed in the results. Table 3 presents the accuracy scores for examiners and raters for both sections of the Simulated Theft experiment. To present the complete data for the second section, the accuracy scores for the last 4 groups were added to the results for the 54 groups. The temporal sequence would place the first 18 groups of the second section immediately after the third 18 groups of the first section.

A word of explanation is in order about the terms "Examiner", "Rater I", and "Rater II." There were five examiners and the accuracy scores refer to the total of the scores for all examiners. The assumption underlying such additivity is that lie detection ability is non-differentiable among them. This assumption is not untenable as later results will show. Furthermore, the overall evaluation of the preliminary training sessions revealed no fundamental differences in technique among the five trainees.

TABLE 3
ACCURACY OF EXAMINERS AND RATERS FOR BOTH SECTIONS OF THE SIMULATED THEFT EXPERIMENT

Subject Groups	FIRST SECTION Accuracy Scores			SECOND SECTION Accuracy Scores		
	Examiner	Rater I	Rater II	Examiner	Rater I	Rater II
1st 18 groups	14	14	17	12	10	12
2nd 18 groups	12	9	8	13	11	11
3rd 18 groups	$\frac{10}{36}$	$\frac{13}{36}$	$\frac{12}{37}$	$\frac{14}{39}$	$\frac{15}{36}$	$\frac{14}{37}$
Last 4 groups Totals for SECOND SECTION				$\frac{4}{43}$	$\frac{4}{40}$	$\frac{4}{41}$

"Raters" refers to the five examiners in their roles as raters, when they were evaluating the records obtained by other examiners. "Rater I" refers to the first set of ratings done soon after the experiments were completed. When both sessions were finished and when every record was rated both by the Examiner and Rater I, another rating was obtained. This was Rater II who was neither the Examiner nor Rater I for that experiment. In other words, each record was rated by three different persons: Examiner, Rater I and Rater II. Only the examiner knew the subjects. The raters evaluated the records independently, never knowing the persons who were involved in the experiments.

In view of the above explanation, it is apparent that the accuracy scores for each column of Table 3 represent the performance of the same people, but under different circumstances. The comparison of the totals of each column is clearly uncontaminated by differences in basic rating ability. The equivalence of these totals would indicate that the different conditions under which the records were rated produced no significant effect on rating accuracy.

Table 3 indicates several striking facts. In the first place, the examiners seem to have benefitted very little in accuracy from observing the subject, testing him, and then analyzing his records. The raters did equally well without this direct contact with the subject. The accuracy scores for the first section were remarkably consistent (36, 36, 37); those for the second section were only slightly less consistent (39, 36, 37).

As for the overall accuracy, that for the first section was practically equivalent to that found for the second. True, there was a slight increase for the second section and this was due entirely to the examiners. But this was not the increase expected after an experiment has been improved and after the examiners (and raters) have had the experience of testing and rating 54 earlier groups. In addition, both examiners and raters had the opportunity of reviewing the errors they had committed in the first section before they began experimenting with the groups in the second.

There is no ready explanation for the lack of substantial increase in accuracy from the first to the second section. It is possible that, though the skills of the examiners (and raters) were much improved, the character of the subjects may have deteriorated. Thus, a selection factor may have operated in providing the less reactive subjects for the second section. Or, word may have gotten around despite the pleas for cooperation and threats of later testing for collusion, and the subjects in the second section may have been more sophisticated, more knowledgeable, and therefore either less reactive or more equally reactive, whatever role they were assigned in the experiments. But there has been no definite evidence to substantiate these conjectures. One is tempted, then, to entertain the notion that there may be some upper bound in accuracy for experiments such as these and that such a bound may have been approached by the examiners and raters.

The division of the groups into successive thirds revealed no significant or meaningful trends. Thus, of six successive groups, the best and the worst accuracy scores (examiner and rater scores combined) are found in the first and second sections respectively.

2. Significance of Results

Some indication of the statistical significance of the results can be presented in the form of Chi-square tables. These are collected in Table 4 and the examiners' data are used for illustration. The results are highly significant, with the Chi-square values far beyond those found in the usual tables. The square roots of the corrected Chi-square values give the estimated standard normal deviates. These range from 9.67 (i.e. $\sqrt{93.6}$) to 11.57 (i.e. $\sqrt{133.82}$) with associated tail probabilities of the order 2×10^{-22}.

There was no doubt, then, that experimental lie detection procedures led to correct discriminations far exceeding chance expectation. However, the importance of this conclusion stemmed from its

TABLE 4
CHI-SQUARE TABLES AND CHI-SQUARE VALUES (X^2)
FOR EXAMINERS' DATA

	1st SECTION (54 Groups)			2nd SECTION (54 Groups)			2nd SECTION (All 58 Groups)		
	f_o	f_e	X^2	f_o	f_e	X^2	f_o	f_e	X^2
All Correct	36	9	81.0	39	9	100	43	9.7	114.9
Not All Correct	18	45	18.2	15	45	20	15	48.3	23.0
	54	54	99.2	54	54	120	58	58.0	137.9
Corrected for continuity			93.6			116			133.8

applicability to examiners and raters alike, since the experimental results were practically equivalent for both. And this meant that lie detection was not primarily dependent upon the intuitive observational skills of the examiner. The physical reactions of the subjects themselves provided indications of involvement or non-involvement in a hypothetical crime and these were clear enough for raters to make accurate discriminations.

3. Agreement Among Raters

Secondary to the problem of accuracy was that of agreement among the various raters. If there was consistency among the raters, then this would imply that either the records were clearly differentiable or that the training of the raters was uniform. But if both accuracy and consistency were high then one would have to conclude that the records provided the objective indices to make precise discriminations. Table 5 presents the agreement scores, analogues of accuracy scores, between examiners and raters and between the raters. Agreement scores, then, refer to the number of groups for which both raters come to exactly the same decision, that is, the number of groups in which each subject was given identically the same designation (T, L or I) by each rater irrespective of the correctness or incorrectness of the decision. In studying Table 5, one is first impressed by the uniformity in the agreement score totals for the first section. Reference to Table 3 will reveal that the corresponding accuracy scores were practically the same.

As for the agreement scores, one observes that they seemed to be slightly larger in the second section than in the first. In addition, there

TABLE 5
CONSISTENCY BETWEEN EXAMINERS AND RATERS IN TERMS OF FREQUENCY OF COMPLETE AGREEMENTS

Sub-Grouping of Sessions	FIRST SECTION			SECOND SECTION		
	Examiner and Rater I	Examiner and Rater II	Rater I and Rater II	Examiner and Rater I	Examiner and Rater II	Rater I and Rater II
1st 18 Experiments	14	17	13	14	12	15
2nd 18 Experiments	9	8	9	13	12	14
3rd 18 Experiments	13/36	12/37	13/35	14/41	14/38	14/43

was less variability among the scores. One might expect this on two accounts. In the first place, the experiment was changed slightly in order that the discrimination among the three roles--Thief, Lookout, Innocent Suspect--could be made with greater precision. Secondly, the raters held an evaluative meeting, after the first section was completed, in order to discuss the errors they had made in their decisions. Both of these situations would tend to improve agreement and produce more homogeneity among the sub-groupings.

One final comment would seem appropriate for the second sub-group of the first section. This group of 18 experiments (54 subjects) yielded the lowest accuracy scores in Table 3 and the lowest agreement scores in Table 5. An examination of the protocol records and research notes gave no clues as to possible experimental or selection variation that would affect this sub-group more than the others.

4. "Serious" Errors

Of all errors possible in the lie detection situation none is more serious than that imputing guilt to an innocent individual. If an error were to be committed, one would generally prefer to free a guilty person than to jail an innocent suspect for a particular crime. The Simulated Theft experiment provided very limited data to explore the tendency of examiner or rater to make these errors. Despite their paucity, the errors of calling the Innocent Suspect a Thief (I → T) are collected together in Table 6. For comparison the "Thief-into-

Innocent" (T → I) errors are also included.

TABLE 6
THE INNOCENT-THIEF (I → T) AND THE THIEF-INNOCENT (T → I) ERRORS FOR THE SIMULATED THEFT EXPERIMENT

ERROR	FIRST SECTION			SECOND SECTION			TOTAL
	Exam	Rater I	Rater II	Exam	Rater I	Rater II	
I → T	1	2	2	0	1	2	9
T → I	4/5	3/5	4/7	1/1	2/3	1/3	15/24

The first impression gathered from the table was the relative infrequency of both the Innocent-Thief and the Thief-Innocent errors. Furthermore, these errors tended to be fewer in the second section despite the fact that it contained 58 experiments, four more than the first section. In general, too, the T → I errors are more numerous than the I → T errors, although there is an equalization of both for the second section. Finally, the examiners made only one of the "serious" errors (I → T) in the 112 experiments of both sections combined.

The reduction of these, the more serious errors, from the first to the second sections could be explained in part by the improvement of methodology that was introduced after the first section. The greater number of T → I errors, as contrasted with the I → T errors, was not unexpected since the Thief was instructed not to reveal to the examiner, either by behavior or by the test response, that he was involved in the theft. In other words, the Thieves were to behave as if they were innocent. The reverse was not expected for the Innocent Suspects.

B. DIFFERENCES AMONG EXAMINERS

An overall analysis, such as that in the previous section, tends to obscure many specific factors that deserve individual attention and evaluation. One of these concerns differences in accuracy among examiners. Accuracy in an examiner may mean one of two things: being accurate in evaluating his own records, or being accurate in evaluating the records obtained by other examiners. If great differences in accuracy exist, one would like to know whether these are due to intellectual, motivational, training, or personality factors. In view of their

equivalent educational level at the beginning of the project, one could argue in favor of homogeneity of intellectual attainments among the examiners. Similarly, there was no evidence that they differed in motivation. Their training was uniform and their progress during the training period showed no outstanding differences. There remains, then, the vague factor of personality as a pertinent variable. Although personality variations were not extreme, appreciable differences were apparent among the examiners, especially in their reactions to and in their impact upon people. If this variable were an important factor, its effects should be most pronounced in the examiner accuracy scores, for this role, in contrast to that of the rater, involves an interpersonal encounter with the suspect who is tested.

1. Accuracy of Examiners

Each examiner evaluated his records and these, in turn, were rated by two of his co-workers. Consequently, the accuracy of each examiner could be compared with the accuracy of other raters on the same set of records. Correlatively, since the examiner evaluated both his own records and those of his colleagues, the accuracy of the examiner under each of these conditions was available for comparison.

In the discussion that follows, the data will be presented in terms of percentage accuracy, that is, the percent of the total number of subjects correctly classified as to the role they played in the Simulated Theft. This statistic is different from that used earlier, i.e., the total number of experiments correctly evaluated. Table 7 presents the

TABLE 7
PERCENT ACCURACY OF EXAMINERS IN EACH OF THEIR DUAL ROLES (AS EXAMINERS AND AS RATERS)

EXAMINER	ROLE				COMPARISON DATA	
	As Examiner		As Rater		Other Raters	
	(N)	(%)	(N)	(%)	(N)	(%)
B	75	75	195	80	150	77
C	108	78	177	73	216	76
D	60	73	30	80	120	82
E	69	87	243	81	138	75
F	24	92	39	85	48	83
Weighted Average		79		78		78

equivalent educational level at the beginning of the project, one could argue in favor of homogeneity of intellectual attainments among the examiners. Similarly, there was no evidence that they differed in motivation. Their training was uniform and their progress during the training period showed no outstanding differences. There remains, then, the vague factor of personality as a pertinent variable. Although personality variations were not extreme, appreciable differences were apparent among the examiners, especially in their reactions to and in their impact upon people. If this variable were an important factor, its effects should be most pronounced in the examiner accuracy scores, for this role, in contrast to that of the rater, involves an interpersonal encounter with the suspect who is tested.

1. Accuracy of Examiners

Each examiner evaluated his records and these, in turn, were rated by two of his co-workers. Consequently, the accuracy of each examiner could be compared with the accuracy of other raters on the same set of records. Correlatively, since the examiner evaluated both his own records and those of his colleagues, the accuracy of the examiner under each of these conditions was available for comparison.

In the discussion that follows, the data will be presented in terms of percentage accuracy, that is, the percent of the total number of subjects correctly classified as to the role they played in the Simulated Theft. This statistic is different from that used earlier, i.e., the total number of experiments correctly evaluated. Table 7 presents the

TABLE 7
PERCENT ACCURACY OF EXAMINERS IN EACH OF THEIR DUAL ROLES (AS EXAMINERS AND AS RATERS)

EXAMINER	ROLE				COMPARISON DATA	
	As Examiner		As Rater		Other Raters	
	(N)	(%)	(N)	(%)	(N)	(%)
B	75	75	195	80	150	77
C	108	78	177	73	216	76
D	60	73	30	80	120	82
E	69	87	243	81	138	75
F	24	92	39	85	48	83
Weighted Average		79		78		78

percent accuracy scores for each examiner both in his role as examiner and as rater. Comparison data have been included to indicate the accuracy of the other raters using the same records that have been evaluated by the examiner. The number of cases for which each percentage has been computed is also presented.

Since there were two raters for each examiner on any one record, the number of cases for "Other Raters" was twice that for the "As Examiner" column. The last percent column, then, is an average of the percent accuracy for the two raters who evaluated the records for a particular examiner.

Table 7 reveals an appreciable spread in accuracy among the examiners, from 92 for F to 73 for D. Despite the magnitude of this extreme difference, the Chi-square test yields a value of 2.4, indicating that these two examiners need not be considered significantly different in their ability to identify correctly the role of each subject in the Simulated Theft experiment. Neither were there any other significant differences in accuracy among the examiners in their roles as examiners.

But how does the examiner perform as a rater? If the interpersonal contact of examiner and suspect in a lie detection examination is helpful in evaluating the guilt-innocence status of the suspect, then the accuracy percent in the examiner role should exceed that for the rater role. Table 7 provides no evidence for this hypothesis. Some of the examiners perform better as examiners than they do as raters. The reverse is true of others. In fact the data suggest a well known phenomenon--the regression effect. The better examiners decrease in accuracy as raters while the poorer examiners improve in their rater scores. In terms of averages, the percent accuracy for the examiner role is 79 and that for the rater role is 78, a very small decrease in accuracy when the evaluation is done on the basis of the records alone.

Although of limited generality, an additional observation can be made concerning the accuracy data of the examiners in their two-fold role. There is some tendency for the better examiner to have a better score as rater. Such a positive relation between the two roles is indicated by a Spearman rank coefficient of .68. There was some hesitation in making this observation because of the inherent instability of a statistic based on so few cases. It was made, nevertheless, because this is a type of result that would be expected among a large group of examiners. Furthermore, it is not likely that an experiment of sufficient scope to test this hypothesis will ever be performed.

The structure of this experiment also permitted the comparison of the accuracy scores of examiner and independent rater in analyzing

the same set of records. This is done by comparing the percentages in the "As Examiner" and "Other Raters" columns. Thus, examiner B obtained an accuracy percent of 75, but the average accuracy for the two raters who evaluated the same records was slightly higher, namely, 77. Although the comparisons may appear interesting, there is a decided limitation to this type of analysis because of the positive relationship between the accuracy scores for the two roles of examiner and rater. In view of this relationship, it is rather obvious that the best examiner will probably have two raters whose evaluation abilities are inferior to those of the examiner. Similarly, the examiner with the lowest accuracy score will most likely have his records rated by those who are more accurate in their evaluations.

There is one final analysis that is pertinent to the results presented in Table 7. It is perhaps the most important in this section. It would answer the question: Were the individual examiners operating well beyond the chance level? The answer is that every examiner attained an accuracy far exceeding chance expectation. In this type of experiment, where the respective roles of three persons were to be identified by the examiner (or rater), the probability of the correct identification of a single individual was one-third. Thus, in evaluating the roles of 30 subjects (i.e. in 10 experiments), ten of them would be accurately identified by chance, that is, 33 percent accuracy. The accuracy level of each examiner was well beyond this point. As an illustrative case the significance test associated with the lowest accuracy figure, namely, 73 percent for examiner D, yielded a Chi-square value of 41.41875 which, in turn, was equivalent to a standard normal deviate of 6.4358 with an associated probability of the order of 2×10^{-10}. It should be clear that the accuracy percents for the other examiners would be even more significant (F excepted because of fewness of experiments).

What is particularly gratifying in the data of Table 7 is that the accuracy in a complicated lie detection experiment can reach a figure close to or beyond 90 percent for some examiners. The lie detector operators had only about 20 minutes examination time and effectively two records on which to base their judgments. Furthermore, they had to make a decision on each case. They were not permitted to use the designation "Inconclusive", thus evaluating only the records for which they had the most confidence. Neither did they have the opportunity of re-examining suspects for hours on end, as is often the case in actual lie detection examinations. The accuracy figure for each examiner in this experiment is highly significant both from a statistical and practical point of view, when one considers the limited sample of behavior that was evaluated.

2. Agreement Among Examiners

A relatively high degree of accuracy for each examiner implies a high degree of agreement among them. In the design of this experiment there were two general types of agreement scores: (a) the agreement of the examiner and the raters (examiner-rater); and (b) the agreement between the two raters (rater-rater). Consequently, depending upon the function he had in a particular experiment, the operator could be involved in an examiner-rater situation if he were the examiner and his evaluations were compared with those of a rater; in a rater-examiner situation, if he were a rater and his evaluations were compared with those of the examiner; in a rater-rater situation, if he as a rater were compared with another rater. The first of the paired roles indicates the person's function in the experiment; the second, the function of the other person involved in the comparison.

It was hypothesized that the rater-rater agreement scores would be higher than the examiner-rater and the rater-examiner scores. It was felt that the latter two comparisons would involve a contamination of two types of lie detection criteria, namely those based on the chart records alone and those based upon the behavior of the suspect during the lie detection examination. Table 8 presents the three types of agreement scores for each examiner. Specifically, for B, the percent agreement score of 73 indicates that 73 percent of his identifications

TABLE 8
THREE TYPES OF PERCENT AGREEMENT SCORES FOR EACH EXAMINER

	TYPE OF AGREEMENT		
EXAMINER	Examiner-Rater	Rater-Examiner	Rater-Rater
B	73	77	76
C	77	80	81
D	83	67	93
E	78	79	81
F	83	80	85
Weighted Average	78	78	81

agreed with those of the two raters who evaluated his records. The score of 77 indicates that when B rated the records of the other examiners he agreed with their designations 77 percent of the time.

Finally, the score of 76 represents the percent of agreement between B and another rater when both were assigned as raters to the same examiner's records.

The most characteristic feature of the data in Table 8 is the uniformly larger rater-rater agreement score for each rater, there being but one inversion in the scores, that of examiner B, who had a score of 77 for the rater-examiner agreement and a score of 76 for the rater-rater agreement. Since a rater-rater agreement score is one of three for each examiner, the probability of its being the largest is one-third, assuming equal probability for each score to be maximal. The probability that four or more examiners have a maximal rater-rater agreement score is slightly less than .05. If the .05 level of significance is used, then it might be concluded that two raters, working with the chart records alone, tend to agree more closely in their ratings than do examiners and raters. This is a reasonable conclusion in view of the possibly different criteria used by examiners and raters. The examiners, it may be recalled, had conducted the lie detector test and may have been affected in their evaluations by the behavior of the suspects.

However reasonable such a conclusion may be, it must be tempered by the very small differences among the various agreement scores for each examiner. Specific reference is made to the data of examiners B, C, and E, whose scores are based on sample sizes of well over 100 and in the greater proportion of cases on sample sizes in the neighborhood of 200. (The agreement scores for examiners D and F, especially those in the second and third columns, are based on fewer than 40 scores.) The weighted averages also indicate a relatively small superiority for the rater-rater agreement scores, a superiority which does not attain statistical significance. Consequently, one can conclude that the agreement between two raters is consistently better than that between an examiner and a rater, but that the degree of improvement is very small and not significantly greater.

The above conclusion tends to confirm the result that the effect of personal contact with the suspect has but a minor influence upon the objectivity of an examiner's evaluation of his records. The earlier result stressed the accuracy of decisions, and the conclusion was that an independent rater attained an accuracy in his identification of the suspects as high as that attained by the examiner who tested them. The results of this section emphasize that an examiner and a rater agree in their evaluations almost as closely as do two raters. From both sets of conclusions one may draw a rather strong implication as to the necessity of having the examiner evaluate the records he obtains. Since the raters are as accurate as the examiners, their evaluations can be

used in place of the examiners'. And if the rating procedure is objective, one can devise a program for a computer, which in turn could be used to displace the raters. Thus, the input to the computer would be exactly the same as that received by the raters who have a chart record of the electrical variations reflecting the physiological changes in the suspect.

Theoretically, then, the problem of an objective, impersonal, machine decision poses no difficulty. In fact, machine analysis of the physiological patterns can be more accurate than the visual inspection procedures generally used in lie detection work. Other criteria, which only harmonic analyzers could adequately identify and measure, may also be developed. All these considerations would argue for the development of computerized procedures in the evaluation of guilt or innocence. Such theoretical implications are valid and inevitable. However, grave practical considerations should delay the immediate application of computer techniques to the physiological indices now in general use among lie detection operators. These practical issues are more dramatically formulated as questions. Are these the best systems for lie detection work? To what extent are these systems under control of the suspect? To what extent can these systems be manipulated by the suspect so as to invalidate the objective criteria of deception the rater or computer uses? It should be apparent that the best computer would fail completely in its lie detection capabilities if its decisions could be manipulated or controlled by the suspect.

C. EXAMINERS IN BOTH SECTIONS OF EXPERIMENT

1. Accuracy Data

Several minor changes were introduced into the experiment after the completion of the first 60 groups. Since these were expected to improve accuracy, a comparison of the data in both sections should reveal whether any significant changes had occurred, specifically in the examiners' accuracy scores. Table 9 presents the accuracy percent for all examiners and for both sections of the experiment. By an accuracy percent score is meant the percent of all subjects correctly identified by the examiner as Thief, Lookout, or Innocent Suspect. Thus, of the 30 subjects examined by examiner B in the first section, 22 were correctly identified for a percent accuracy score of 73.

Of major importance in Table 9 are the data of examiners B, C and E since the data for examiner F are based on relatively few cases and those for examiner D are incomplete. With this in mind, one notices but one major change in accuracy from the first section to the second. Examiner C improved appreciably in his role as examiner during the second session.

TABLE 9
EXAMINER ACCURACY PERCENT SCORES IN BOTH SECTIONS OF THE SIMULATED THEFT EXPERIMENT

EXAMINER	When Examiner				When Rater			
	1st Section		2nd Section		1st Section		2nd Section	
	N	%	N	%	N	%	N	%
B	30	73	45	76	111	79	84	81
C	60	70	48	88*	60	73	117	73
D	30	73	30	73	30	80	--	--
E	36	86	33	88	102	77	141	83
F	6	100	18	88	33	82	6	100
Weighted Percent		76		82		78		79

*Significant increase at the .01 level.

In general, the examiners improved in accuracy in both roles as they progressed from the first to the second section of the experiment. However, the accuracy increases were small, as demonstrated by the insignificant rises in the weighted percents. In fact, the weighted accuracy percents were substantially uniform for both sections of the experiments and for both roles assumed by the examiner. Thus, whether he evaluated the records of a subject he personally examined or whether he evaluated a set of records of a subject completely unknown to him, the examiner attained effectively the same accuracy.

Specifically, the small gains in accuracy from the first to the second section do not reflect the anticipated improvements expected as a result of the changes introduced into the experimental procedure.

2. Consistency Data

After the completion of the first section of the experiment, the examiners held a review session among themselves to check on their procedures and to discuss the types of errors they made. It was anticipated that such group self-evaluation would result in greater homogeneity in their subsequent ratings. This should be reflected in greater consistency among raters who had to evaluate the same set of records.

In the design of this experiment an examiner's consistency with other raters could be analyzed in three different ways. Table 10 presents

the percent agreement scores for these three rating situations. Although the data of examiner F are presented in the table, they are placed in parentheses to indicate that the scores are based on too few cases and consequently do not possess stability. Examiner D did not participate in the rating procedures of the second section, and there are no data for that portion of his results.

The outstanding feature of the data in Table 10 is the overall significant rise in the agreement scores from first to second section.

TABLE 10
PERCENT AGREEMENT SCORES FOR THREE RATING CONDITIONS IN BOTH SECTIONS OF THE SIMULATED THEFT EXPERIMENT

EXAMINER	RATING CONDITIONS					
	Examiner-Rater		Rater-Examiner		Rater-Rater	
	1st Section	2nd Section	1st Section	2nd Section	1st Section	2nd Section
B	70	76	70	87**	68	86**
C	71	85*	80	79	72	85*
D	73	93**	67	--	93	--
E	75	83	75	81	73	89**
F	(100)	(78)	(76)	(100)	(82)	(100)
Weighted Percent	73	83*	75	82*	72	87**

** Significant at .01 level. *Significant at .05 level.

With one exception for examiner C, each agreement score increases for the four examiners whose data possess sufficient stability for individual analyses. The review session among the examiners, following the first section of the Simulated Theft experiment, undoubtedly had a pronounced effect on the agreement among the ratings. Although this review of rating procedures and errors was aimed at increasing accuracy, its primary effect seems to have been an increase in consistency.

Finally, it may be pointed out that the rater-rater consistency scores were the largest among the three rating conditions. This was anticipated. Largest, too, was the gain in consistency within this rating

situation. The review session induced greater uniformity in the rating procedures, and this uniformity guaranteed greater inter-rater consistency without improving accuracy to a corresponding degree.

D. ROLE DIFFERENTIATION--DISCRIMINANT ANALYSIS

The critical task of a lie detection operator is role differentiation. He has to decide, for example, whether the suspect is a criminal or whether he is innocent. Since the polygraph records include data from several physiological response systems, the problem of determining the relative importance of each system is of paramount importance. Whether the operator is completely aware of such differential weighting is difficult to say. But his final decision, insofar as it is independent of any clues obtained from the behavior of the suspect, must reflect some weighting scheme.

In this experiment, the examiners (and raters) had ample opportunity to develop such procedures. They were required to score each critical response for each physiological system, recording these data on one sheet. This enabled the examiner to obtain a comparative picture of the relative variations in the different scores before he arrived at his decision. In view of its greater response variability and the relative simplicity in its interpretation, the psychogalvanic reaction was considered by the raters as probably most influential in their final decisions. (In practice, lie detector operators undoubtedly differ in the importance they attach to the various physiological reactions. Some emphasize blood pressure changes; some favor respiratory criteria; and some depend upon the psychogalvanic reaction more or less exclusively. The manuals give lip service to integrating the "whole record", but they are not clear as to how one may achieve this in practice.)

The discussion, up to this point, has been centered upon the examiner and the rater who, consciously or unconsciously, use some weighting procedures to aid them in arriving at a decision. It was pointed out that such procedures were subjective and relatively unique, varying from person to person. To eliminate the subjective element from such weighting schemes, special discriminant scores were developed and computed from the data gathered in this experiment. There were three of them: the purpose of the Thief-Innocent score was to differentiate between the Thief and the Innocent Suspect; the Lookout-Innocent and the Thief-Lookout scores had similar functions. Such scores were computed for each physiological index, the larger score indicating greater manifest involvement in the crime. In this way, the basis was formed for a discriminant function which was used to render an objective determination of the roles assumed by the subjects.

1. A Typical Set of Data. Discriminant Scores

To clarify the discussion which is to follow, the T-I (Thief-Innocent) discriminant scores for the data of Examiner D are presented in Table 11. Examiner D tested 10 experimental groups, the numbers of which are shown in the first column. Thus, in Experiment 67 the Thief obtained a T-I score of 7 for the respiratory

TABLE 11

T-I (THIEF-INNOCENT) DISCRIMINANT SCORES FOR THE PHYSIOLOGICAL INDICES, THE TOTAL, AND THE DISCRIMINANT FUNCTION (TYPICAL SET OF DATA)

Exp No	THIEF (T)					INNOCENT SUSPECT (I)				
	X	Y	Z	T	D	X	Y	Z	T	D
67	7	5	21	33	52	5	4	10	19	23
73	5	6	18	29	51	0	0	4	4	12
74	1	7	11	19	43	2	1	10	13	26
79	3	7	3	13	14	7	4	7	18	8
84	4	4	14	22	38	9	2	4	15	-10
91	0	6	12	18	47	6	7	4	17	8
94	14	12	26	52	60	6	9	9	24	27
102	3	12	13	28	53	3	9	4	16	20
110	4	10	14	28	49	5	9	7	21	24
120	7	12	16	35	51	2	6	9	17	33
	4.5	9	9	9	10	(Accuracy in discrimination)				

X: Respiratory T-I Score
Y: Plethysmographic T-I Score
Z: Psychogalvanic T-I Score

$T = X + Y + Z$
$D = 100 (-.029X + .019Y + .030Z)$
Discriminant Function Score

responses, 5 for the plethysmographic reactions, and 21 for the PGR. The sum of these was 33, yielding a Total score which, in effect, gave equal weights to each of the physiological indices. On the other hand, the T-I scores for the Innocent Suspect in Experiment 67 were 5, 4, and 10 respectively, yielding a Total of 19. In this particular experiment, each physiological index demonstrated the existence of greater emotional disturbance (i.e. larger T-I scores) in the Thief. It would have been possible, then, to make a correct diagnosis on the basis of any of the physiological scores, although the differentiation appears much clearer for the PGR.

The results of Experiment 79 did not follow the same pattern. Here, two of the three physiological indices yielded T-I scores which

would have led to a decision clearly at variance with the facts, since the Innocent Suspect would have been designated as the Thief.

An examination of the remaining experiments points up the need for some overall score which would exhibit more stability than those associated with the individual physiological indices. Although the Total score is one such possibility, its use would imply that all three indices are equally important in predicting the Thief-Innocent status of the subjects, and this is not the case. The results, as summarized at the bottom of the table, tend to show that for this set of data the respiratory index (4.5) is no better than chance in classifying the subjects into the Thief and Innocent Suspect categories. The other two physiological indices classify nine of the ten pairs correctly. Similarly, the Total score, used as an index, attains the same degree of accuracy, namely, nine out of ten correct classifications. But the Discriminant Function scores, computed for this set of data, attain perfect accuracy with 10 correct classifications.

The original data, in the form of chart recordings for these 10 experiments, were also evaluated by two independent raters. To complete the analysis, two additional tables like Table 11 were constructed. The accuracy scores of the three physiological indices were computed, as well as those for the Total score and the Discriminant Function score. This was done for the data of all examiners. There were two restrictions which made it impossible to utilize all 112 experiments. For each examiner, only those experiments were included which were scored by the same two raters. The data for Examiner F were not used because the sample was small---six cases. These requirements eliminated 19 experiments from this analysis.

The example in Table 11 illustrates only one phase of the investigation. Lookout-Innocent (L-I) and Thief-Lookout (T-L) discriminant scores were also subjected to similar analyses. The next portion of this section will be concerned with the relative accuracy of the physiological indices and the relative efficiencies of the weighted indices in classifying the subjects of the experiment.

2. Classification Accuracy of the Various Indices

The data in Table 11 suggested several important comparisons. For example, the relative accuracies of the three physiological indices could be assessed. Moreover, the Total (T) and the Discriminant Function (D) scores, as examples of weighted indices, could be compared. To emphasize the differences existing among them, the data for the three discriminations, namely, T-I, L-I, and T-L, could be analyzed.

Finally, the results from the first section of the experiment (Experiments 1-60) could be contrasted with those in the second section. Table 12 presents the first of two tables that point up the meaningful

TABLE 12

THE RELATIVE ACCURACIES (AS PERCENTAGES) OF THE SEVERAL SCORES FOR THE THIEF-INNOCENT, LOOKOUT-INNOCENT, AND THIEF-LOOKOUT CLASSIFICATIONS (DATA FROM ENTIRE EXPERIMENT)

DISCRIMINATION	SOURCE OF SCORES	DISCRIMINANT SCORES					
		N	X	Y	Z	T	D
Thief-Innocent	Examiners	93	67	71	90	90	97
	Raters	198	56**	69	89	88	93*
Lookout-Innocent	Examiners	93	61	72	85	85	90
	Raters	198	60	71	84	84	87*
Thief-Lookout	Examiners	93	58**	54**	76	70	80
	Raters	198	54**	59	72	68	79*

X: Respiratory Score
Y: Plethysmographic Score
Z: Psychogalvanic Score
* Based on 186 sets of ratings

$T = X + Y + Z$
$D = 100 (w_x X + w_y Y + w_z Z)$
where w_x is the weight for X, etc.
** Not significant

comparisons. As can be observed, the results are kept separate for examiners and raters. Despite the overall agreement, the percentage accuracy for the examiners' data are very slightly, but uniformly higher than that for the raters' data. Thus, in the Thief-Innocent discrimination, for the respiratory scores, the accuracy for the examiners was 67 percent, for the raters 56 percent. These differences are in favor of the examiners' data in all but one instance, the plethysmographic score in the Thief-Lookout discrimination, where the raters' accuracy was 59 percent compared to 54 percent for the examiners.

Several points may need emphasis at this time. The examiners and the raters did not make these discriminations. Their rated data, transformed into discriminant scores, were evaluated for accuracy in discriminating between all possible pairings of the roles assumed by the subjects in this experiment. Further, the Total (T) scores and the Discriminant Function (D) scores are merely linear combinations of the same three physiological scores. The difference between the two consists in the arbitrariness with which the T scores were obtained, and the uniqueness

of the D scores. In other words, the adding of the three physiological scores is an arbitrary assignment of a weight of + 1 for each. Another person might have chosen, for example, weights of 1, 2, 1 for the X, Y, and Z scores; or, the weights 1, 2, 3; or, for that matter, any set that he felt reflected the importance of the three physiological responses. On the other hand, the weights in the discriminant function, being the only linear set which maximizes the accuracy of classification, are unique.

The next point to be mentioned is the way in which the Discriminant Function scores were used. In many problems, one Discriminant Function score is computed for the averages of the variables in one classification, and another for the averages in the other classification. In this way, the midpoint of these two Discriminant Function scores constitutes the cut-off point, serving to classify any particular individual as above or below it and, therefore, as belonging to one classification rather than the other. To illustrate the procedure, a D score would be computed for the (M_x, M_y, M_z) vector for the Thief data, another for the (M_x, M_y, M_z) vector for the Innocent Suspect data, and these would be the D scores for the "average" Thief and for the "average" Innocent Suspect. The midpoint of these would serve as the cut-off score, the criterion for separating individuals into the Thief and into the Innocent Suspect categories. But in the present investigation, each Thief was paired with each Innocent Suspect in a unique manner, since only these two could be validly compared in a particular experimental session. Consequently, the D scores for these paired individuals were compared directly and no "average" Thief or "average" Innocent Suspect was used as a reference base for classification purposes.

Therefore, the percentage accuracy figures in Table 12 indicate how efficiently the rated data itself can discriminate between the two individuals, one of whom is known to differ from the other in a very significant way. Consequently, the accuracy percentages indicate the efficiency in discriminating between such pairs as the Thief-Innocent, the Lookout-Innocent, and the Thief-Lookout. This is a much easier task than that which often confronts the lie detector operator in real life. He may have to examine only one individual and, consequently, would have no other records providing meaningful comparisons. Or, if he has to examine two or more individuals, there is no guarantee that one or more of them are actually implicated in the matter under investigation. In contrast, the task involved in the discriminant analyses of this experiment is a relatively simple one. It is, however, a necessary first step in the appraisal of the efficiency of lie detection procedures.

In returning to Table 12, one is impressed with the differences in accuracy among the three physiological scores. Thus, uniformly

the psychogalvanic scores have higher accuracy (when compared with the other two physiological measures) for all three types of discriminations, the Thief-Innocent, the Lookout-Innocent, and the Thief-Lookout. In every instance, they are at least 13 percent units higher. Of the three physiological measures, the ratings on the respiratory responses generally produce the least accurate discriminations.

Of the two weighted scores, T and D, it is clearly apparent that better accuracy is obtained from the discriminant function which generated the D values. This is an important result. For, whatever may be meant by a "global" or "overall" analysis of the records, the examiner (or rater) would rarely be able to arrive at the best possible weighting scheme in his visual evaluation of a set of records. For the inexperienced operator, the mental task of differentiating responses within a physiological mode is difficult enough. To have him mentally assign differential values to the modes themselves complicates his task to such an extent that he may fall victim to doing the easiest thing, ---just counting the number of "large" reactions wherever he finds them. And this is an equal-weighting procedure, which is not as efficient as that determined by discriminant function analysis.

The most "unusual" result in Table 12 appears in the comparison of the accuracy figures for the psychogalvanic scores (Z scores) and the Total scores. Here, a single score is better than the sum of three scores, each of which presumably serves as an indicator of the same phenomenon---"lying reaction". This should serve a warning against the indiscriminate combination into one score of several predictive indices. Such practices are often analogous to adding adulterants to a good product. The process of properly combining several different indicators is not merely a matter of adding several scores. It is a relatively complex process of determining the relative value of each component for doing the job at hand most efficiently.

The discriminants used in this research were the simple best fitting linear functions. Radical departures from linearity would require more complicated analyses. In these experiments the simple linear discriminant, as exemplified in the D scores, was the most efficient index among those considered.

The last important comparison is that among the different pairings of the experimental roles. From an examination of the accuracy percents for the Z, T, and D scores, it appears that the most efficient discrimination was between the Thief and the Innocent Suspect; and that the differentiation between the Lookout and the Innocent Suspect was more efficient than that between the Thief and the Lookout. This coincides with expectation, and it also agrees with the feelings of the raters, who found greatest difficulty in differentiating the Thief and the

Lookout. This well defined result for the Z, T, and D scores is not so sharply delineated in the X and Y scores (the pneumograph and the plethysmograph). Though the Thief-Lookout discriminations are lowest in accuracy, the differentiation in efficiency between the Thief-Innocent and the Lookout-Innocent discriminations is all but obliterated for these two physiological indices.

One final comment is in order. It is indeed heartening to observe that if one takes the examiners' data and uses these in a discriminant function to decide which of two individuals is the Thief and which the Innocent Suspect, the impersonal mathematical function will select the Thief correctly with an accuracy of 97 percent. In a similar fashion, the mathematical function will make a correct discrimination between the Lookout and Innocent Suspect in 9 out of 10 decisions. Finally, even in the difficult Thief-Lookout discrimination, the accuracy for the examiner's data reaches an accuracy of 80 percent, a figure well above chance expectation. (For the examiners' data, significance at the .05 level is reached when the accuracy is about 60 percent; for the raters' data, the significant percentage is about 57.)

The accuracy percentages for the mathematical discriminant functions reflect both the validity of the physiological measures and the accuracy of the ratings. The most unusual feature in these results is the relatively high accuracy for the Lookout-Innocent discrimination. This discrimination was based on the responses to only two of the critical questions. The other striking feature is the 97 percent accuracy in the Thief-Innocent discrimination, a figure averaged from the data of four examiners. In general, then, the data of well trained examiners tend to be highly and consistently valid in differentiating between two such roles as Thief and Innocent Suspect.

3. Accuracy for First and Second Sections of the Experiment

Since, midway in the experiment, slight changes were made in the procedure, the data from the first and the second sections were analyzed separately in order to see if any significant improvement could be observed in the accuracy figures for the various indices. Table 13 presents the results for both sections and for examiners and raters.

A general overview of the results indicates that the accuracy percentage figures for examiners and raters are relatively homogeneous. Whether he be an examiner, or a rater of another's records, the operator evaluates the physiological responses in a consistent fashion. As was noted in the previous section, the mathematical discriminant function (D) yields greatest accuracy. The Z and T accuracy figures follow

closely behind, while the X and Y scores are least efficient. Furthermore, the Thief-Innocent discrimination is most accurate and that for the Thief-Lookout least accurate.

The important feature of Table 13 lies in the comparison between the first and second sections of the experiment. As had been mentioned

TABLE 13

THE RELATIVE ACCURACIES (AS PERCENTAGES) OF THE SEVERAL DISCRIMINANT SCORES FOR THE FIRST AND SECOND SECTIONS OF THE EXPERIMENT

DISCRIMINATION	SECTION	SOURCE OF SCORES	N	DISCRIMINANT SCORES X	Y	Z	T	D
Thief-Innocent	First	Examiners	43	71	59	93	93	98
		Raters	86	62	63	89	85	90
	Second	Examiners	50	63	81	88	88	96
		Raters	112	51	73	89	89	96
Lookout-Innocent	First	Examiners	43	57	66	81	89	84
		Raters	86	59	63	81	82	83
	Second	Examiners	50	65	77	88	89	96
		Raters	112	61	77	87	85	90
Thief-Lookout	First	Examiners	43	59	43	73	67	79
		Raters	86	60	49	67	58	71
	Second	Examiners	50	56	64	78	72	80
		Raters	112	49	66	75	75	86

in Chapter II, the changes in the experiment were expected to make the Lookout role more clearly differentiable from the other two. The results seem to fulfill this expectation. To illustrate, there is no appreciable change in accuracy between the first and second sections for the Thief-Innocent discrimination. However, the data for the Lookout-Innocent and the Thief-Lookout discriminations clearly show that greater accuracy was achieved for the second section of the experiment. Thus, for the X scores (Lookout-Innocent) the examiners' data yielded an accuracy percentage of 65 for the second section and 57 for the first section. The corresponding percentages for the raters were 61 and 59. In fact, all of the percentage figures for the second section are higher than those for the first section when the L-I discrimination is the one under

consideration. In the Thief-Lookout discriminations, only the X scores fail to show greater accuracy for the second section. The data, then, reflect the changes in the experiment much more clearly in this analysis than they do in the overall accuracy figures (p. 57) that were obtained from the examiners' and raters' decisions as to the roles played by the subjects in each experimental group.

4. The Discriminant Functions: Invariance Problem

Earlier in this section, the discriminant function for the data in Table 11 was found to be: $-.029X + .019Y + .030Z$. This expression indicates that the best weighted index was obtained by multiplying the respiratory, plethysmographic and psychogalvanic scores by weights that bear the corresponding ratios of -3:2:3. The discriminant functions of two other raters, who evaluated the same charts were: $-.013X + .0076Y + .023Z$ and $-.027X + .012Y + .017Z$. Although these expressions do not seem to be directly comparable, division by the coefficient of Z in each case would facilitate the analysis of the homogeneity of the three discriminant functions derived from the same data by the three different raters. In this particular set of discriminant functions, the relative ratios of the coefficients (multipliers) are:

	X:	Y:	Z
Examiner	-.97:	.63:	1.00
Rater I	-1.59:	.71:	1.00
Rater II	-.57:	.33:	1.00

In all instances, it is seen that the X score (respiratory index) has a negative weight, whereas the Y score (plethysmographic index) has a positive weight, though smaller than that obtained by Z (psychogalvanic index) which is taken as a reference base.

The results presented above would have been very promising, if they could have been duplicated for the other examiners and raters. One would have had some confidence that a reasonably invariant function could be developed from the ratings of the physiological responses. And if this were the case, one would have established both the reliability and objectivity of the ratings within the limits of a scaling factor. Unfortunately, this was not the case as the results in Table 14 clearly

demonstrate. Thus, the data of examiner B show that both the X (respiratory index) score and the Y (plethysmographic index) score would be multiplied by 1.55 and the Z (psychogalvanic) score by 1.00 to yield the most accurate Thief-Innocent discriminations for his data. On the other hand, the same charts when evaluated by Rater I

TABLE 14
RELATIVE RATIOS OF THE THIEF-INNOCENT DISCRIMINANT WEIGHTS FOR ALL EXAMINERS AND RATERS IN THE SECOND SECTION OF THE EXPERIMENT

EXAM	Examiner			Rater I			Rater II		
	X	Y	Z	X	Y	Z	X	Y	Z
B	1.55	1.55	1.00	-.30	1.46	1.00	.46	.46	1.00
C	-2.55	-.68	1.00	-.36	.33	1.00	.00	.38	1.00
D	-.97	.63	1.00	-1.59	.71	1.00	-.57	.33	1.00
E	.43	1.06	1.00	.54	.33	1.00	-.28	-.00	1.00
F	-1.05	1.27	1.00	.58	.14	1.00	.33	.05	1.00

yielded a discriminant function that would require multipliers in the ratios of -.30: 1.46: 1.00 to yield the most accurate Thief-Innocent discrimination. When the same physiological responses were rated by Rater II, the multipliers were .46: .46: 1.00. The data of examiner C produced even greater variability in the proportions. In fact, for any examiner and the two raters who scored the same records, there is little evidence for homogeneity among the ratios between the multipliers.

If there is no appreciable equivalence in the discriminant weights among those who rated the same records, one would not expect much consistency in the discriminant weights among examiners who rated different sets of records. An examination of the first third of Table 14 shows no strong consistency patterns among the examiners. Thus, the ratios obtained from the data of examiner C show no similarity to the ratios for the other examiners. Examiners B and E have all positive weights, while D and F have negative X-weights, the others remaining positive. There does not seem to be much homogeneity, then, among the relative ratios of the discriminant weights from examiner to examiner, or from examiner to rater, even when the latter evaluates the same set of records as the examiner.

But, perhaps there is some consistency in the discriminant patterns for the same operator---that is, when he evaluates the records at one time as an examiner and at another as a rater. With this as a possible hypothesis, the data of those examiners who served several times as raters (for other examiners) were collected and are summarized in Table 15. As the table indicates, examiner B served as rater for two other examiners. Examiners C and E were raters for four different examiners. Among these data, two types of consistency may be considered: that between the roles of examiner and rater; or that among the several instances in which the role was the same, namely rater. The table demonstrates no appreciable homogeneity in the examiner-rater discriminant ratios. Neither is there any substantial equivalence manifest a ong the ratios for the several instances in which the examiner served as rater.

It is quite apparent that the search for invariant discriminant functions may be discouraging insofar as the present data are concerned. However, it is just as apparent that no serious hope of consistency

TABLE 15
RELATIVE RATIOS OF THE THIEF-INNOCENT DISCRIMINANT WEIGHTS OF THREE EXAMINERS WHO SERVED AS RATERS FOR SEVERAL OTHER EXAMINERS

EXAM	RELATIVE RATIOS OF WEIGHTS When Examiner			When Rater		
	X	Y	Z	X	Y	Z
B	1.55	1.55	1.00	-.36	.33	1.00
				.54	.33	1.00
C	-2.55	-.68	1.00	-.30	1.46	1.00
				.58	.14	1.00
				-.57	.33	1.00
				-.28	-.00	1.00
E	.43	1.06	1.00	-1.59	.71	1.00
				.46	.46	1.00
				.00	.38	1.00
				.33	.05	1.00

should have been entertained, considering the small sizes of the samples

from which the discriminants were computed. But when the same data were rated by several experienced operators, one could not help but expect greater uniformity in the discriminants. That this did not happen indicates rather strongly that any hope for computerization in the lie detection field must be based on instrumental measurements and not on human ratings. Once the measurements are made with complete objectivity, rater variability will be eliminated and only sampling variability will remain to affect the discriminant function. In view of the fundamental problems which are yet to be solved in the objective measurement of the physiological patterns that may be useful in lie detection, the concern for obtaining optimal discriminant functions is a small matter indeed.

Only the Thief-Innocent data were used for illustration in this section. It was felt that this data afforded the greatest opportunity of demonstrating the possible invariance of the discriminant function, at least in those instances where the same data were evaluated by three different raters. Also, the analysis was confined to the second section of the experiment in which better controls were available. Despite this selection of the data, the results did not offer any evidence that invariance would be easy to attain. The lack of homogeneity of the results clearly indicated that, before any advances could be made in this direction, rater variability was the first factor that had to be eliminated, however inconsequential it may have appeared in the analysis of the records for accuracy.

E. ASSOCIATION AND PEAK-OF-TENSION TESTS

Three critical words, all related to the Simulated Theft, were used in a free association test to provide additional evidence for differentiating the Innocent Suspect from the Thief and from the Lookout. These were: PAMPHLET, LOOKOUT, COINBOX, in the first section; and BUTTON, LOOKOUT, COINBOX in the second section. The words, in a fairly direct manner, reminded the Thief and the Lookout of their intimate involvement in the "robbery." Consequently, these stimuli should have induced greater excitement in the Thief and Lookout than in the Innocent Suspect. The words BUTTON and COINBOX should have helped in discriminating between the Thief and the Lookout.

In the modified Peak-of-Tension test, the Thief was required to incriminate himself by replying, to a number called by the examiner, with another number which represented the exact amount of money stolen. This single number response was expected to differentiate between the Thief and the Innocent Suspect (T-I), and between the Thief and the Lookout (T-L) as well. No differential reaction was expected between the Lookout and Innocent Suspect (L-I).

In evaluating the efficiency with which each word discriminated among the members of each experimental group, the Total score, a combination of the scores obtained on all three physiological response systems, was employed. To provide greater reliability, the Total scores for all three raters were used in the final analysis.

In order to facilitate the evaluation, the results were kept separate for the first and second sections of the experiment. Furthermore, each word and the critical number (in the modified Peak-of-Tension test) were used in a three-fold analysis. Thus, each stimulus word (or number) was evaluated for its efficiency in discriminating between Thief and Innocent Suspect, Lookout and Innocent Suspect, and Thief and Lookout. To test the significance of the discriminations, a median test was used for each critical word (or number) and a corrected Chi-square value was computed for each of the three pairings of subject roles. These results are presented in Table 16 for both sections of the experiment.

As can be noted in the table, the word PAMPHLET failed to produce discriminating physiological responses between any of the paired roles. It served its purpose, however, since it was included as a contrast to the other association stimuli. It was the only word which was used in the directions to all subjects and served to focus in the minds of all of them the object (pamphlet rack) associated with the "crime." The Innocent Suspects, then, were clearly alerted to associate the word PAMPHLET with the "crime" and with the specific purpose of the lie detector test. Under these circumstances, it was not surprising that this word manifested no discriminating power. It produced equally disturbing associations in all three classes of subjects.

On the other hand, the situation was quite different with the stimulus word BUTTON. It was most unlikely that the Innocent Suspect would associate this word with the "theft." Theoretically, only the Thief knew of its existence. The Chi-square value of 6.81, significant at the .01 level, indicated that the Thief was more excited than the Innocent Suspect while giving an associated response to the word BUTTON. There was no difference between the reactions of the Lookout and Innocent Suspect (Chi-square value, .03), and this was what one would expect. As for the Thief-Lookout discrimination, the Chi-square value of 3.45, though close to the critical value of 3.84, was not significant at the level required in this experiment.

The stimulus word LOOKOUT was quite effective in discriminating between the Lookout and Innocent Suspect, as indicated by the significant

TABLE 16
CHI-SQUARE VALUES FOR CRITICAL WORDS AND NUMBERS DENOTING THE EFFICIENCY OF DISCRIMINATION FOR THE T-L, L-I, AND T-L PAIRINGS OF ROLES IN THE FIRST AND SECOND SECTIONS OF THE EXPERIMENT

	FIRST SECTION				SECOND SECTION		
STIMULUS	ROLE DISCRIMINATIONS			STIMULUS	ROLE DISCRIMINATIONS		
	T-I	L-I	T-L		T-I	L-I	T-L
Pamphlet	.33	.59	.91	Button	6.81**	.03	3.45
Lookout	.15	7.28**	1.83	Lookout	2.21	4.97**	.86
Coinbox	5.34*	16.42**	.15	Coinbox	9.97**	9.97**	.03
124	13.38**	1.52	10.76**	124	35.41**	6.81**	12.51**

* Significant at .05 level.
** Significant at .01 level.

Chi-square values of 7.28 and 4.97 for the first and second sections, respectively. However, it was anticipated that this word would also be effective in discriminating between the Thief and the Innocent Suspect, but the insignificant Chi-square values failed to verify this expectation in either section of the experiment.

Of the three association stimuli, the word COINBOX consistently performed as hypothesized. In both sections of the experiment, there was significant differentiation in the T-I and L-I discriminations, insofar as the physiological reactions to this word were concerned.

The last stimulus in the table refers to the critical number used in the modified Peak-of-Tension test. The subject was required to say the next higher number, 125. Since the relation of these numbers to money had been emphasized in the directions, the Thief was required to state an amount of money exactly equal to the amount stolen. It was anticipated that this procedure would make two effective discriminations, namely, in the Thief-Innocent, and Thief-Lookout pairings. The results in the table clearly justify this expectation with extremely significant Chi-square values for the two discriminations. But it is not altogether clear why the L-I discrimination yielded a significant result. Two possible explanations

may help explain this outcome. It is not improbable that some Lookouts may have guessed the correct amount that was taken by the Thief from the coinbox. There is also some, but small, likelihood that in a few cases the Thief may have wittingly or unwittingly revealed to the Lookout how much was taken. If the directions were correctly followed, the entire experimental procedure generated considerable excitement with the "robbers," and this state of mind may have brought about some instances where slips could have occurred. This, too, may be a partial explanation for the failure of the BUTTON stimulus to yield significant results for the T-L discrimination in the second section of the experiment.

To summarize briefly: association words are effective to the extent that they are intimately associated with the critical event by the guilty and not by the innocent; Peak-of-Tension stimuli are effective if only the guilty person knows the specific item under consideration while the innocent individual cannot infer its relevance to the crime.

F. INTERCORRELATIONS AMONG THE PHYSIOLOGICAL INDICES

There are no definitive data on the degree to which the three physiological systems intercorrelate when they are used in lie detector examinations. This experiment was planned so that some information on this matter could be gathered.

Antecedent to the discriminant function analysis, the respiratory, plethysmographic, and psychogalvanic indices, illustrated in Table 19, were intercorrelated. These coefficients were computed for the results of each examiner and each rater and for the first and second sections separately. In each case an estimate was obtained of the correlation between each pairing of the physiological indices. The Spearman rank-difference correlation coefficients were used.

Some explanation may be required regarding the variables being correlated. Basically, they were "excitement scores" or "lie scores" determined from the ratings and combined into T-I, L-I, or T-L discriminant scores. Since such scores were kept separate for the Thief, the Lookout, and the Innocent Suspect, the intercorrelations were based on a somewhat homogeneous range of emotional reactivity. Thus, it was anticipated that the T-I scores computed for the Thieves would show less variability than the T-I scores for Thieves and Innocent Suspects combined. The same reasoning holds for the intercorrelations among the three physiological T-L scores, as computed for the Thieves and Lookouts separately; similarly, for the L-I scores.

In effect, then, the correlation coefficient measured the concomitant variations in physiological excitation among Thieves, Lookouts, and Innocent Suspects. For the first two classes of subjects, the excitation was definitely associated with lying. As for the Innocent Suspects, the intercorrelations would reflect the extent of relation among the physiological reactions due to the excitement associated with taking a lie detector test. In view of the different attitudes these groups were required to assume in the experiment, the results will be analyzed separately for each of them. The Fisher transformation was used in averaging the various rank-order correlation coefficients. Table 17 presents the "average" correlation coefficients between the three different pairings of the physiological variables. These coefficients were obtained from the Fisher z-values by an inverse transformation.

TABLE 17

CORRELATION COEFFICIENTS BETWEEN RESPIRATORY, PLETHYSMOGRAPHIC AND PSYCHOGALVANIC DISCRIMINANT SCORES FOR THE THIEF, LOOKOUT AND INNOCENT SUSPECT CATEGORIES IN THE FIRST AND SECOND SECTIONS OF THE SIMULATED THEFT EXPERIMENT

ROLE	VARIABLES CORRELATED	SECTIONS OF EXPERIMENT	
		FIRST	SECOND
THIEF	Resp-Pleth	.31	.07
	Pleth-PGR	.18	.15
	Resp-PGR	.30	.13
LOOKOUT	Resp-Pleth	.30	.22
	Pleth-PGR	.46	.24
	Resp-PGR	.22	.27
INNOCENT SUSPECT	Resp-Pleth	.41	.20
	Pleth-PGR	.36	.30
	Resp-PGR	.24	.25

From an overall point of view, the intercorrelations are low. However, all are positive. In general, the intercorrelations are lower in the second section of the experiment. Finally, if one were to average the corresponding coefficients for both sections of the experiment, it would appear that those among the Thieves would be lowest when compared

with the Lookout and Innocent Suspect coefficients.

It was noted that the intercorrelations for the Innocent Suspects were obtained under conditions of truth-telling whereas those for the Thieves and Lookouts were obtained while these subjects were lying. Despite the differences in emotional attitude due to lying and despite the generalized excitement induced by the roles the Thieves and Lookouts had to play, the intercorrelation matrices do not appear to differ appreciably from one class of subjects to another. The only deviant group seems to be that for the Thieves during the second session of the experiment.

As for low intercorrelations, such a state of affairs is favorable for prediction purposes, provided the correlations with the criterion are appreciable. For it is rather obvious that combining highly interrelated variables adds little towards increasing the efficiency of prediction. In this particular experiment, the discriminant function score was a linear combination of three physiological indices which had low intercorrelations. This score was the most accurate discriminator, that is, the best predictor of the role status assumed by the subjects. However, it did not seem to be very much more accurate than the psychogalvanic score alone. The reason for this was the rather high correlation of the psychogalvanic score and the criterion (i. e. high accuracy of discrimination) and the rather low discriminative accuracy in the respiratory and plethysmographic reactions.

G. OBJECTIVITY OF RATINGS (DISCRIMINANT SCORES)

The discriminant scores, Thief-Innocent (T-I), Lookout-Innocent (L-I), Thief-Lookout (T-L), were computed for the three physiological indices, and within these indices, for each examiner and his two raters. It was thus possible to compute the consistency (objectivity) of ratings for each of the physiological systems by correlating two sets of scores for the same index. For example, the ratings of an examiner and one rater, on the psychogalvanic response, would be correlated. In like manner, the correlation coefficients between the same examiner and the other rater and between the two raters could be calculated. Thus, there were available, as estimates of the objectivity in the ratings of each physiological index, three correlation coefficients for each of the three separate discriminants.

There were many different variables involved in the determination of these coefficients. For example, there were differences due to the pairings of raters who could be either examiner-rater I, examiner-rater II, or rater I-rater II. Since there were three discriminants used, correlation coefficients were computed within each discriminant category, that is, T-I, L-I, T-L. Finally, separate correlation

coefficients were computed for the Thief, Lookout and Innocent groupings.

The main objective, however, was to obtain an estimate of the objectivity of ratings for each of the three physiological indices; respiratory, cardiovascular (plethysmographic), and psychogalvanic reactions. In view of this objective, it was decided to combine the three sets of coefficients for each pair of raters, since it was a reasonable assumption that the raters were equivalent in rating ability. Next, it was assumed that the coefficients for the Thief, the Lookout, and the Innocent Suspect should be kept separate since there was homogeneity of mental attitude within each category but pronounced differences between the categories. As an example, it is quite evident that the mental state of the Thief was very different from that of the Innocent Suspect undergoing the same lie detection examination. Consequently, it was decided to combine the coefficients for the Thief category in the T-I and T-L discriminant. In similar fashion, the Lookout coefficients were combined from the T-L and L-I discriminants. The same procedure yielded a combined Innocent Suspect coefficient. The combination procedure involved transforming each Spearman rank-difference correlation coefficient into a corresponding Fisher z-value. The z-values were averaged by weighting each inversely as its variance. These averages, in turn, were transformed to correlation coefficients. Table 18 presents the coefficients for each of the three physiological indices, for the first and second sections of the experiment, with the results kept separate for the Thief, Lookout, and Innocent Suspect categories.

The results in Table 18 show a good degree of uniformity from the first to the second sections of the experiment. Furthermore, the coefficients for each index among the three roles are no less homogeneous than those between the two sessions of the experiment.

The outstanding result in the table is the low average value of the correlation coefficients obtained from the respiratory and cardiovascular responses. These range in the low .40's for respiration and about .50 for the cardiovascular changes as recorded by the plethysmograph. Contrary to popular claims of the objectivity of such records, these coefficients indicate a relatively low order of consistency among the ratings of the same set of records for these physiological responses. This is due mainly to the complexity of the physiological responses in question, and the disappointing results suggest that a visual analysis, as usually performed in actual lie detection investigations, may be a relatively subjective matter. The low inter-rater coefficients may, in large part, explain why the intercorrelations among the physiological

TABLE 18

"AVERAGE" CORRELATION COEFFICIENTS EXPRESSING DEGREE OF OBJECTIVITY AMONG RATERS IN EVALUATING THE PHYSIOLOGICAL RESPONSES OF THIEF, LOOKOUT, AND INNOCENT SUSPECT FOR THE FIRST AND SECOND SECTIONS OF THE EXPERIMENT

ROLE	PHYSIOLOGICAL INDEX	SECTIONS OF EXPERIMENTS	
		FIRST	SECOND
THIEF	Respiratory	.48	.46
	Cardiovascular	.57	.39
	Psychogalvanic	.81	.80
LOOKOUT	Respiratory	.39	.43
	Cardiovascular	.51	.50
	Psychogalvanic	.83	.83
INNOCENT SUSPECT	Respiratory	.38	.35
	Cardiovascular	.46	.53
	Psychogalvanic	.77	.85

indices were so low (cf. previous section). Furthermore, they may indicate why the discriminating accuracies of these indices were so low throughout this experiment.

On the other hand, the consistency among raters in their evaluation of the psychogalvanic responses is shown by the rather substantial correlation coefficients which range in the low .80's. The magnitude of these coefficients reflects the relatively good discriminative capability of the psychogalvanic response.

As a general overview, the results indicate a low degree of objectivity in the evaluation of the respiratory and plethysmographic reactions by ordinary visual examination of the records. The complex character of these physiological patterns would seem to require more formidable methods of analysis than the "global" or "clinical" evaluations used by lie detector operators.

H. CONFIDENCE IN DECISIONS: EXAMINERS AND RATERS

Each rater and examiner was requested to indicate the degree of confidence he had in his decisions as to who was the Thief, the Lookout,

and the Innocent Suspect. A six point rating scale was used for this purpose (cf. Appendix E).

Despite the slight variability among the raters and examiners, the results were averaged for all raters and examiners for the first and second sections. The ratings for correct identifications were kept separate from those in which errors of diagnosis were made. The results are presented in Table 19. The higher the score, the greater the expressed confidence in the decision.

TABLE 19
AVERAGES OF CONFIDENCE RATINGS FOR CORRECT AND INCORRECT IDENTIFICATIONS OF ROLES OF SUBJECTS FOR FIRST AND SECOND SECTIONS OF EXPERIMENT

SECTION	RATINGS OF	CORRECT IDENTIFICATIONS			INCORRECT IDENTIFICATIONS		
		T	L	I	T	L	I
FIRST	Examiner	3.78	3.54	4.20	2.58	2.58	2.55
	Rater	3.81	3.53	3.86	3.10	3.21	3.54
SECOND	Examiner	3.83	3.63	3.91	2.36	2.67	2.20
	Rater	3.88	3.71	4.03	3.04	3.48	3.26

The outstanding, but not unexpected result was that the raters had greater confidence in those decisions which were ultimately verified as correct than they did in those which were incorrect. To illustrate, the examiners in the first section exhibited an average confidence rating of 3.78 in their correct Thief identifications as compared with the rating of 2.58 for the incorrect identifications of the Thief. This type of contrast was found for all paired comparisons of confidence ratings among correct and incorrect identifications.

An unexpected result was the relative equivalence of the ratings for the first and second sections of the experiment. It was anticipated that the change introduced into the experimental procedure for the second section would have increased the confidence of raters and examiners in their decisions. But no general and uniform increase in confidence was observed from the first to the second section.

As for the differences between examiners and raters, these will be treated separately for the correct and incorrect identifications. An examination of the ratings among the correct identifications revealed a striking equivalence of results for examiners' and raters' estimates of their feelings of confidence. Thus, for the Thief category, the average ratings in the first section were 3.78 and 3.81 for examiners and raters, respectively; for the second session, 3.83 and 3.88. The largest difference was in the ratings of Innocent Suspects. In the first section the examiners were slightly more confident; in the second, the raters. In general, then, the confidence of the examiner was the same whether he made his decisions on his own cases (as examiner) or on the cases of other examiners (as a rater).

Quite a different result is observed for the incorrect identifications. In every instance the rater expresses more confidence in his wrong decisions than does the examiner. Furthermore, these differences are appreciable and consistent from first to second section. It is as if the grave responsibility for making a wrong decision was not so seriously considered by the rater. This was to be expected. The examiner who tested the individuals may have felt greater pressure and greater concern in arriving at a correct decision. The rater, on the other hand, was relieved of considerable anxiety in this regard, since, if he made an error in diagnosis, he could blame the error on inadequate testing by the examiner. It is precisely this implication that the examiner wished to avoid when rating his own records. He could blame no one but himself for an inadequate examination.

The final comparison was concerned with the degree of confidence in ratings of the Thief, the Lookout, and the Innocent Suspect. Although these ratings were not independent, the comparisons among the roles, for the correct identifications, indicate that the Innocent Suspect was rated with greatest confidence and the Lookout with least confidence. The differences are small but they are consistent. This conclusion, however, does not hold for the incorrect identifications among which no clear cut pattern emerged.

In general, then, the results from the confidence ratings indicate that correctly identified roles were associated with greater confidence than those incorrectly identified. Among the correct decisions, there was a trend for the Lookout to be judged with least confidence. In addition, the confidence ratings were equivalent for examiners and raters and between first and second sections of the experiment. The only significant trend observed among the incorrect identifications was the uniformly greater confidence of the raters in their erroneous decisions.

I. INTERPRETABILITY OF THE RECORDS

To estimate the ease or difficulty with which the responses could be interpreted, the three physiological responses were evaluated on a four point rating scale (Appendix E) by the examiners and raters. These ratings were made, for each subject, at the time when the final decision as to the role he played in the Simulated Theft was given. The ratings of all examiners and raters were averaged for the entire experiment since no appreciable differences were observed between the results of the first and second sections. The larger average ratings indicate greater ease in interpreting the responses. The results are presented in Table 20 for the correct and incorrect identifications among examiners and raters.

TABLE 20

AVERAGE RATINGS OF THE INTERPRETABILITY OF THE THREE PHYSIOLOGICAL RESPONSES BY EXAMINERS AND RATERS FOR CORRECT AND INCORRECT IDENTIFICATIONS

SUBJECT ROLE	INDEX RATED	EXAMINERS		RATERS	
		Correct	Incorrect	Correct	Incorrect
THIEF	Resp	2.12	2.30	2.05	2.00
	Pleth	2.20	2.43	2.27	2.22
	PGR	2.74	2.70	2.79	2.29
LOOKOUT	Resp	2.00	2.37	2.15	2.12
	Pleth	2.33	2.59	2.19	2.15
	PGR	2.73	2.81	2.64	2.37
INNOCENT SUSPECT	Resp	2.39	1.86	2.51	2.33
	Pleth	2.43	2.00	2.45	2.37
	PGR	2.73	2.36	2.71	2.60

The most significant and most consistent result in this table indicates that, with but one minor exception, the order in which the physiological indices were rated for ease of interpretability was: psychogalvanic, plethysmographic, and respiratory. In general, the psychogalvanic reaction was rated appreciably easier to interpret than the other two responses; among the correct identifications, it was also rated with greatest consistency among examiners and raters alike.

In one respect, the results in Table 20 run contrary to expectation. One would ordinarily anticipate that the correctly judged records would also have been easier to interpret. Though this expectation was upheld by the results for the raters, this was not the case among the examiners. In fact, six out of nine comparisons among the examiners' data indicated that the incorrect identifications were based on records that were judged as easier to interpret.

The only significant conclusion to be gleaned from these data is that the psychogalvanic response was deemed easiest to interpret among examiners and among raters. This was an expected result in view of the relative simplicity of the psychogalvanic response pattern in comparison to that obtained in respiratory and plethysmographic reactions. The respiratory pattern was judged more difficult to interpret than the plethysmographic reaction, but the differences in these two ratings were very small.

J. UNINTERPRETABLE REACTIONS

The relative complexity of the attachments placed on each subject, along with the stress of the examining procedure itself, led to the anticipation that some reactions would become contaminated by artifacts and consequently be difficult to interpret. It was, for example, impossible to prevent the subject from moving a bit or shifting his position during the examination. These unavoidable occurences were minimized by the instructions, an initial period of self-adjustment, and the relatively short periods of experimentation. Despite these precautions, occasional random and uncontrolled movements occurred which made the interpretation of that part of the record difficult or impossible. In addition, changes in the external environment, such as an occasional loud noise in the corridor outside the testing room, sometimes produced effects in the subject which were not due to the experimental stimuli. Finally, lapses on the part of the examiner either disturbed the subject or produced a response that was not adequately controlled. As specific examples, noise created by unnecessary movements of the examiner or asking a question before the subject returned to his normal level of response, may be cited.

The records of all subjects were examined for reactions that could not be validly interpreted. These were grouped separately for the Thief, the Lookout and the Innocent Suspect. The data were also categorized into correct and incorrect identifications for the first and second sections of the experiment. The results are presented in Table 21. Since the distribution of these uninterpretable responses was appreciably skewed, medians were computed for each of the groups.

TABLE 21
MEDIAN NUMBER OF UNINTERPRETABLE RESPONSES IN THE RECORDS OF THE THIEF, LOOKOUT, AND INNOCENT SUSPECT FOR CORRECTLY AND INCORRECTLY IDENTIFIED RECORDS

ROLE OF SUBJECT	FIRST SECTION		SECOND SECTION	
	Correct	Incorrect	Correct	Incorrect
THIEF	4.5	4	5.5	9
LOOKOUT	5	5	6	6
INNOCENT SUSPECT	3.5	3	4.5	5.5

As a specific example, one may observe that the Lookouts who were correctly identified in the first session had a median number of five uninterpretable reactions in their records. The Lookouts who were incorrectly identified by the examiners also obtained a median score of five. In the second section, the correctly and incorrectly identified Lookouts produced a median number of six uninterpretable responses.

The median scores for the second section are uniformly larger than the corresponding scores in the first section. This is due to the fact that there were 75 critical responses to evaluate in the second section, while there were only 66 such responses in the first section. In both sections, then, the median number was considerably less than 10 percent of the total number of responses available for interpretation. As a matter of fact, in most instances where movements or external stimuli introduced irrelevant reactions in the records, the examiners repeated the procedure and used the new reactions in place of those that were invalid.

The expectation that the records of the Innocent Suspects would be least affected by artifacts seems verified by the low median scores exhibited by these subjects. However, a median test for discriminating the Innocent Suspects from the other two groups did not yield significant results in any of the four columns of the table. But if the results for the entire table are considered, the probability that the Innocent Suspects would obtain the smallest median score in each of four comparisons (with the Thief and Lookout groups) would be of the order of .012, on the assumption of equal likelihood for each group to be the lowest of the three.

The important comparison, however, is that between the correctly and incorrectly identified groups. It was conjectured that a possible reason for incorrect decisions was the presence of artifacts in the records. Thus, it was argued that there are nervous or tense subjects who, despite the roles they played, might reflect this tension by producing records which would be more difficult to interpret. On this assumption, such "difficult" subjects would tend to produce more uninterpretable reactions in their records. This hypothesis, however, was not verified by the data. The subjects who were incorrectly identified, when compared with those who were correctly identified, did not produce a significantly larger number of uninterpretable reactions. In fact, there are two instances in which the correctly identified subjects obtained a larger median score (4.5 vs 4 and 3.5 vs 3). In the case of the Lookouts, for example, there was no difference between these two groups. Even the correctly and incorrectly identified Thief groups in the second section do not differ significantly from each other, despite the relatively large difference between the medians (9-5.5).

However, the uninterpretable scores also contained reactions produced by uncontrolled stimuli which were external to the subject, such as, an unusual noise. As explanations for incorrect decisions, these cannot be considered as important as for example, movements originating in the subject being tested. In the case of movements, these could be interpreted as arising from tension, or as being deliberately induced to deceive the examiner. With this in mind, the number of movements were analyzed for significance in differentiating the correctly and incorrectly identified subjects. Because these responses were relatively infrequent, with a median response of one movement per subject, there was insufficient variability in median scores from group to group. At any rate, it was relatively clear that the correctly and incorrectly identified subjects did not differ appreciably for any of the groups (Thief, Lookout, Innocent Suspect) in either the first or second section of the experiment. There was, however, a slight tendency for the Thief category in the second section to generate more movements than was found in the records of the other two groups. The summary of these results may be found in Table 22.

K. TENSION-ANXIETY IN THE SUBJECTS

The behavior of the subjects led the supervisors in this experiment to the conclusion that the Simulated Theft engendered considerable tension and anxiety. Feelings of "guilt" and discomfort were also reported by the Thief or the Lookout in a number of groups. These conclusions were based on observation or the volunteered statements of the subjects. A more systematic verification of the impact of this experiment was available in the Tension-Anxiety Rating Scales which were filled out by each subject.

TABLE 22
MEDIAN NUMBER OF MOVEMENTS LEADING TO UNINTERPRETABLE RESPONSES FOR CORRECTLY AND INCORRECTLY IDENTIFIED SUBJECTS IN THE FIRST AND SECOND SECTIONS

ROLE OF SUBJECT	FIRST SECTION		SECOND SECTION	
	Correct	Incorrect	Correct	Incorrect
THIEF	1	1	2.5	2
LOOKOUT	1	1	1	1
INNOCENT SUSPECT	.5	1	1	1.5

1. During the Theft

Both the Thief and the Lookout rated the degree of anxiety each felt during the commission of the "theft." The differences in their anxiety ratings were tested for statistical significance. In both sections of the experiment, the Thieves reported greater anxiety than the Lookouts. The t ratios were 4.00 and 2.94 for the first and second sections, both values significant at the .01 level. This result was expected and served to confirm the observations of the supervisors.

2. While Waiting for the Test

While the subjects were waiting to be tested it was anticipated that the difference in anxiety between the Thief and the Lookout would be reduced. It was not altogether clear what to expect from the Innocent Suspects. Since it was common knowledge that a flurry of small thefts had been a matter of concern of the University administration, it was anticipated that some Innocent Suspects would become suspicious that this was not an experiment but a real investigation. Despite this possibility, it was felt that the Innocent Suspects would report less anxiety than the Thieves. This, in general, was verified, but as the reported data indicate, the results were not statistically significant in the second section of the experiment.

3. During the Testing

During the actual lie detector test, it was conjectured that the suppressed anxiety in the Lookout would erupt to a degree almost equal to that of the Thief, who had already experienced appreciable anxiety in

"stealing" the money and worrying about getting caught. It was assumed that the Thief would have become adapted and that he would have calmed down a bit. Although some anxiety was expected in the Innocent Suspect

TABLE 23
MEAN ANXIETY RATINGS OF SUBJECTS WHILE WAITING FOR THE LIE DETECTOR TEST

SECTION	THIEF	LOOKOUT	INNOCENT	F-RATIO
First	7.02	5.07	5.11	9.80**
Second	5.43	5.46	5.00	.65

** Significant at .01 level.

when actually confronted by a serious and mature interrogator, it was felt that his anxiety would be lowest, and only slightly higher than that he exhibited while waiting for the lie detector test.

An analysis of variance for both sections of the experiment verified what was expected regarding the Innocent Suspects. The difference between the Thieves and Lookouts in self-reported anxiety was greater in the first section than had been anticipated. In general, however, the results are consistent with the hypotheses.

TABLE 24
MEAN ANXIETY RATINGS OF SUBJECTS DURING THE LIE DETECTOR TEST

SECTION	THIEF	LOOKOUT	INNOCENT	F-RATIO
First	6.94	5.96	5.54	4.62*
Second	6.34	6.43	5.10	4.77*

* Significant at .05 level.

4. Feeling of "Giving Self Away"

All subjects were asked to report on how excited they were in responding to the critical questions. Specifically, they were asked to estimate the extent to which their inner physiological reactions have been interpreted as lie responses. The hypothesis was that while answering the critical questions, both Thief and Lookout would feel

more physiological reverberations than the Innocent Suspect.

An analysis of variance of these "guilt reaction" ratings shows pronounced consistency for the first and second sections of the experiment. The F-ratios indicate highly significant results along the lines indicated by the hypothesis. It is quite apparent that the significant outcome is due almost entirely to the low "guilt reaction" ratings of the Innocent Suspects.

TABLE 25
MEAN "GUILT REACTION" RATINGS OF SUBJECTS
DURING THE LIE DETECTOR TEST

SECTION	THIEF	LOOKOUT	INNOCENT	F-RATIO
First	3.48	3.17	1.72	44.34***
Second	3.38	3.17	1.50	53.14***

*** Significant beyond .001 level.

5. Conclusion

The results of the rating scales indicate varying degrees of anxiety experienced by the three types of subjects used in this experiment. In general, the results agree with hypothesis and expectation for both sections of the experiment. The self-report data and the observations of the supervisors tend to corroborate the impression that the task may have aroused anxiety in a manner analogous to that expected of guilty and innocent suspects in real life situations. Whether role expectations influenced the ratings or whether there were genuine "guilt" emotions could not be ascertained in this design.

L. SUMMARY

A number of important results have been established in the Simulated Theft experiment. The first substantiates the confidence of many practicing operators in their instruments and methods of interpretating the polygraph records. In other words, the use of the lie detector in differentiating between the Thief, the Lookout, and the Innocent Suspect in a simulated theft has produced significant results. More important than this general result is the finding that all of the examiners, trained by an experienced lie detector operator, were able to attain a significantly high degree of accuracy in identifying the Thief, the Lookout, and the Innocent Suspect. Lie detection ability, then, is not a unique and innate capacity which an individual possesses or does not possess; rather, it is a skill which can be

learned by a reasonably intelligent person.

Equally important is the result that relatively the same degree of accuracy was attained by those who performed the lie detector test and by those who evaluated only the records. This implies that the lie detector operator was not helped in his diagnostic task by his personal contact with the subjects in the testing situation. This means that the chart recordings of the physiological reactions obtained in an interrogation session possess inherent validity of a degree sufficient to enable an independent rater to make accurate discriminations among the roles assumed by the subjects in this experiment. As a first step, these results were encouraging and positive indications of the feasibility of computerization in lie detection tests.

An additional analysis of the rated responses involved the construction of discriminant scores in order to determine the efficacy, in the lie detection process, of each of the single physiological indicators. For this purpose three types of discriminations were tested, namely, the Thief as differentiated from the Innocent Suspect, the Thief as differentiated from the Lookout, and the Lookout as differentiated from the Innocent Suspect. As anticipated, for each physiological index the Thief-Innocent and the Lookout-Innocent discriminations were made with much greater accuracy than the Thief-Lookout discrimination. In all these differentiations, the psychogalvanic response attained the highest degree of efficiency; and, in general, the respiratory index was least effective.

Since the lie detector operator utilizes the information available in all three physiological systems, linear discriminant functions were computed in order to maximize decision accuracy by assigning the most appropriate weights to each of the physiological indices. With the use of these discriminant functions, uniformly greater accuracy was obtained in each of the basic discriminations between Thief and Innocent Suspect, Lookout and Innocent Suspect, and Thief and Lookout. In fact, for all raters, the percentage accuracy in differentiating the Thief from the Innocent Suspect was 97. Since only the critical questions were used in these analyses, this was indeed a promising result.

The final test of objectivity in the ratings was the comparison of the discriminant functions among the raters and examiners. If these functions were found to be equivalent, the computerization of the lie detection process would be a relatively straightforward matter, since homogeneous discriminant functions would imply objectivity in the rating procedures of examiners and raters. The results, however, did not indicate sufficient equivalence among the discriminant functions,

even among raters who evaluated the same set of records. Though this appears to be a discouraging outcome, its proper interpretation should be based on the fact that limitations in the size of the sample had no little effect in producing unstable estimators.

A brief association test and a modified Peak-of-Tension test were also included in the experimental procedure. Both tests exhibited significant discriminating power and demonstrated the value of using stimuli which had specific and intimate reference to only one person among the several being compared.

Since each physiological response was given a separate rating for each critical question (or stimulus), an estimate of the relationships among the indices, as well as an estimate of the objectivity of the ratings for each index, was available. The inter-index coefficients were low, the majority ranging from .20 to .40. The coefficients expressing the degree of agreement between the ratings of two raters for the same physiological index (objectivity of ratings) were relatively low for the respiratory and plethysmographic indices. They were appreciably higher for those of the psychogalvanic response. The low objectivity coefficients were, in large part, responsible for the low accuracy scores attained by the respiratory and plethysmographic discriminant scores.

The rating scales used with the examiners indicated that a greater degree of confidence was expressed in decisions ultimately shown to be correct than in decisions later found to be wrong. Furthermore, the examiners and raters found the psychogalvanic response easiest to interpret. In a further analysis aimed at clarifying the reasons for incorrect identifications of subjects' roles, it was found that movements and other artifacts found in the records were not significantly more numerous in the erroneous decisions.

Finally, the rating scales, filled out by the subjects upon completion of the experiment, indicated that the simulated theft and the testing situation generated sufficient tension and anxiety in the Thief and Lookout to make these two differ significantly from the Innocent Suspect.

CHAPTER IV

DENIAL-OF-ACTUAL-CRIME EXPERIMENT

A. RATIONALE

The primary objective of this experiment was to provide a more realistic situation for evaluating the lie detection procedure. For this purpose, a group of criminal offenders on parole and in the process of rehabilitation were chosen as subjects. It was assumed that they would be well motivated to deny their criminal background, especially to people who had no authority to obtain such information. Instructions further emphasized this need for self protection and urged them to say NO to any questions that might be related to their former offenses. Since this was described as an experiment to study how various groups reacted to embarrassing questions, they were prepared for any type of disturbing stimuli. They were assured that the examiner had no knowledge of their criminal involvement and had no right to such information.

This experiment, then, did not involve simulation. In denying any question that might reveal his criminal past, the subject was defending himself against unfavorable social biases and strengthening his personality in order to live as a respected member of the community, the very purpose for which he was being prepared in the rehabilitation process.

B. SUBJECTS

Twenty three men, ranging in age from 20 to 35, volunteered as subjects. They were undergoing individual or group therapy at a rehabilitation clinic, to which they were referred by court order because of criminal infractions of the law such as drug addiction, sexual deviation, and larceny. Table 26 presents the offenses and their frequency in this group.

Although they knew they would be paid for their time, a large number of clients refused to participate in the experiment. Even when encouraged by their therapist, they refused. This is not unusual for this type of population since many of its members would try to avoid any interrogation about their private attitudes and activities. Consequently, those who were tested constitute a subsample of volunteers who were probably not overly concerned about answering embarrassing questions.

TABLE 26
TYPES OF CRIMINAL OFFENSES OF SUBJECTS
AT REHABILITATION CLINIC

OFFENSE	FREQUENCY
Drugs	5
Larceny	10
Sex Deviation	4
Other	4
TOTAL	23

C. PROCEDURE

1. Test Questionnaires

Two questionnaires were used in this experiment. These were similar in structure to those used with the college subjects. Because it had been planned to include four critical questions in each form, these questionnaires were longer, each containing 40 questions. Of the four critical questions, two were concerned with the criminal offense for which the subject was brought to court. The other two questions concerned criminal offenses of which the subject was never accused, nor likely to have ever committed. They were selected to be equivalent in seriousness to the actual offenses. Thus, in answering NO to all four critical questions, the subject would be lying on two occasions and telling the truth on the other two. Two of the critical questions, regarding one actual and one fictitious offense, were repeated in the second questionnaire.

Each set of questionnaires was unique in that the critical questions referred to a specific individual. The critical questions were formulated by the supervisor who, with the permission and help of the therapist, extracted the relevant information from the clinical files on the volunteer subject. The critical questions were then incorporated into the questionnaires which were given to the examiner before the testing began. The examiner did not know which of the critical questions pertained to the actual offenses and which to the fictitious offenses. Except for the critical questions, the others were identical for all subjects.

The structure of this experiment required the examiner to detect the two (of the four) questions to which the subject lied. The task was quite

different in the Simulated Theft experiment. There the examiner had to detect the Thief and differentiate him from the Lookout and from the Innocent Suspect. The differentiation in this experiment was between questions and not between persons.

A copy of the skeleton questionnaires (critical questions omitted) may be found in Appendix B.

2. Test Sessions and Instructions

Each subject was interviewed by the supervisor who provided an explanation of the research project prior to the subject's entering the testing room. (See supervisor's instructions in Appendix B.) The supervisor then introduced the subject to the examiner and gave the examiner the two questionnaires to be used for that subject.

The examiner gave the subject the necessary instructions and attached him to the apparatus. (See examiner's instructions to subjects in Appendix B.) A short intermission of 3-4 minutes between both questionnaires served to reduce adaptation effects. After completing the test, the examiner referred the subject to the supervisor to be paid and to receive the final instructions. These exacted a promise from the subjects not to reveal the details of the procedure to anyone.

Appreciable heterogeneity, in mental ability and educational level, was noted among the members of the sample. When, early in the experiment, one or two subjects failed to say NO to questions about their criminal involvement, the supervisor enlarged the directions so as to repeat and stress what had to be denied. The results of those who misunderstood the directions were necessarily excluded from analysis.

3. Test Environment

Insofar as time and accessibility were concerned, the subjects would have experienced great inconvenience to travel to Fordham University. Consequently, the polygraph was shipped to the clinic. A small examining room, approximately 8' x 10', was made available for testing purposes.

Testing began about 4 PM and continued for about five hours, depending upon the scheduled appointments. This continued for two successive weeks, each containing three or four testing days.

The test room was in a reasonably quiet section of the clinic facilities. The weather, however, was relatively uncomfortable at

times, since the testing was done in July and the clinic was not air-conditioned.

4. Examiners

The 23 subjects were tested by three examiners who rotated their duties so that each had approximately equal testing loads. These examiners also alternated as supervisors, one examiner and one supervisor being used in the testing of each subject. Of the five examiners who served in the Simulated Theft experiment, these three had the greatest amount of experience.

5. Evaluating the Records

The three examiners evaluated the critical responses of all 23 subjects on a rating form similar to that used before. After rating each of the physiological indices separately, the raters were required to pick out, from among four critical questions, the two questions on which each subject lied. As a control, two other raters, working only with the charts and completely unaware of the nature of the questions that were used, were given the same records to evaluate. In this "blind" analysis, these raters were also requested to select the two questions on which the subjects lied. For this procedure, the questions were identified merely by number, not by content.

D. RESULTS AND DISCUSSION

In this experiment the rater was required to select, from four questions, the two questions on which the subject lied. Consequently, for any one subject, the rater could get a score of two correct, one correct, or zero correct choices. The frequencies obtained for these three classifications are presented for each of the raters used in this experiment. These are given in Table 27 which also includes the results of tests of significance. Two questionnaires were used and the data are tabulated for each separately (Test Session I and II). Of the 23 subjects, one did not have his records rated in Test Session II because of an invalidating circumstance.

The Chi-squared values were obtained from a two-category table where one category was "Two Correct" and the other "One or None Correct." This analysis was done for each rater separately and for two groups of raters. The basis for grouping the raters was experience. Raters H and S were lie detector trainees who were unaware of the details of this experiment and who rated the records without knowing the nature of the questions that had been used. Raters B, C, and E had

TABLE 27
DISTRIBUTION OF CORRECT CHOICES FOR THE FIVE RATERS AND THE SIGNIFICANCE OF THE RESULTS FOR BOTH SESSIONS

TEST SESSION	RATER	Number Correct: Two	One	None	Chi-Squared
I	B	10	12	1	11.90**
	C	10	8	5	11.90*
	E	8	13	2	5.44*
	Total	28	33	8	28.40**
	H	3	17	3	NS
	S	6	14	3	NS
	Total	9	31	6	NS
II	B	8	13	1	6.14*
	C	6	11	5	NS
	E	5	14	3	NS
	Total	19	38	9	6.99**
	H	4	16	2	NS
	S	5	14	3	NS
	Total	9	30	5	NS

** Significant at .01 level.
* Significant at .05 level.
NS Not Significant

completed their training as operators; in addition, they had considerable experience as examiners and raters in the Simulated Theft experiment. Raters B, C, and E were also the examiners in this experiment and knew the entire experimental procedure. On both accounts, then, their results were expected to show a greater number of correct choices.

To facilitate the understanding of Table 27, rater B's results will be briefly described. In Test Session I, he was able to identify both lie questions in 10 of the 23 subjects. One of the two lie questions was selected in 12 of the 23 subjects. In the case of one subject, neither of the two lie questions was identified correctly. His evaluations were significantly "better than chance" (Chi-squared value of 11.90) since the expected number of two correct identifications is 23/6 a value much smaller than the observed number of 10. In Test Session II, rater B again was able to detect the lie questions with accuracy significantly better than chance.

The results further indicate that the two groups differed in their accuracy, the experienced group obtaining the better scores. This, however, is more characteristic of the first test session, where all three raters in the experienced group performed significantly above chance while neither of the two others attained significant results. In fact, the accuracy scores of these two raters are effectively the same for both sessions, while those for the experienced raters drop sharply from the first to the second session.

Several factors must be kept in mind in evaluating these results. The experienced raters were also the examiners and supervisors in the experiment and, therefore, were well acquainted with the structure of the critical questions and the types of crimes committed by the subjects. Some carry-over effects of this knowledge may have contributed to their higher accuracy scores despite the fact that they rated the records in a relatively objective manner. The two raters, H and S, had no knowledge whatsoever of the types of questions used. Though greater knowledge of procedure and questions may have contributed to the better results found among raters B, C, and E, the obvious factor of greater experience must also have proved effective. Raters H and S were in the middle stage of their training in lie detection and did not yet have any experience as examiners in any phase of this research.

The decrease in accuracy for the second test session would seem to imply that, after the first questionnaire, the subjects adapted rapidly. The anxiety, shame, and tension may have diminished to the point where the subject showed little reactivity to his lies in the second questionnaire. It may be an important fact that individuals who have been subjected to much questionning, as may be the case with criminals, show initial emotional reactivity which may diminish rapidly with continued interrogation.

But what can be said about the value of these results in general? In one sense the outcome of this experiment is disappointing, since one would prefer higher accuracy for the experienced raters. The best score for any rater was 10 of 23 cases in which both lies were correctly identified. As a contrast, the examiners and raters were able to attain complete accuracy in 36 out of 54 experimental groups in the first section of the Simulated Theft experiment. Moreover, the accuracy improved in the second section. Since the probability of detecting both lies among four questions is the same as the probability of correctly differentiating among the Thief, the Lookout, and the Innocent Suspect, it is quite obvious that much greater accuracy was obtained in the Simulated Theft experiment. Despite identical probability expectations, it is likely that the examiner can more easily discriminate among truthful and lying suspects than he can differentiate truth from falsehood in a single suspect.

It was anticipated that there would be a relatively large disturbance manifested by the rehabilitation patient when he would be questioned about his former misdeeds. This may well be the case. But the denial of crimes in which he had not been involved may have produced equally large reactions. In the first place, these "other" crimes may have been perpetrated by his friends or associates and he may have felt that this seemingly innocent experiment was actually an investigation. Along the same lines, the patient may have thought that the examiner was really an interrogator who suspected him of being involved in those "other" breaches of the law. Under these circumstances, such "other" crimes became highly emotional issues in contrast to the crimes to which he had already confessed and which he was denying in an experimental and not in a real life situation.

On the other hand, there may be quite a different explanation for the lack of accuracy in differentiating a lie about a crime the patient had committed from the truthful denial of an "equivalent" crime in which he had not been involved. These patients have admitted their offenses many times, --to the police, to the court, and to the therapists. This was "old stuff," and with repetition brought about lesser and lesser emotional involvement. Therapy, too, helped reduce the former tensions and guilt feelings associated with the transgressions. In contrast to the general adaptation to this familiar type of stimulus, a patient would have become relatively sensitized to questions that would seem to implicate him in other crimes. This, then, would explain the lack of sufficiently accurate discrimination between reactions to lies about one's already admitted crime and reactions to questions about participation in crimes not actually committed by the patient.

It is also possible that volunteer bias may have been a substantial factor in the disappointing accuracy produced in this experiment. Many subjects refused to volunteer and those who did may have had little concern or anxiety in being questioned about their former criminal activity. Again, by contrast, these patients might have shown greater emotional reactivity to being questioned about "other" crimes.

The possibility that the subject may have been involved in the "other" crimes, despite the care with which these were chosen, can not be completely discounted. It is very difficult to be certain that a person did not commit an unwitnessed activity, since there can be no absolute verification except the word of the person himself. It was the denial of "other" crimes, and not the criminal activity for which they were apprehended, that brought about intense physiological

disturbances in two of the patients. Though such reactions can be expected to occur because of special sensitivity to certain matters, the consistency of the reaction leaves open the possibility of involvement, by members of this sample, in other unsolved crimes.

Finally, differentiating the lie from the truth, within a single individual, is undoubtedly a different task from distinguishing a lying person from a truthful one. In the case of the single individual, fewer comparisons were available for evaluation. In the Simulated Theft experiment, both intra-individual and inter-individual comparisons entered into the final decision.

E. CONCLUSION

Though the examiners could differentiate between the lying and truthful responses of former criminals now undergoing therapy, the accuracy with which this was done is not sufficiently high for developing objective criteria that might be available for computer analysis. The objective scoring of the same records by two lie detector trainees did not differentiate significantly between the lying and truthful responses of these subjects. Even two of the three experienced raters did not score above chance expectation with the second questionnaire.

A discussion of the possible explanations for these disappointing results points up the inherent difficulties of working with this type of population in lie detection studies.

CHAPTER V

DENIAL-OF-CLASSIFIED-INFORMATION EXPERIMENT

A. RATIONALE

The purpose of this experiment was to provide another realistic situation with which to evaluate the efficacy of lie detection procedures. Basically, the situation involved interrogation of subjects about "classified" information. So that security regulations would not be violated, the "classified" information was fictitious, but the subjects were not aware of this fact. When tested by the examiner, the subjects denied possessing this "classified" information because they were obligated to deny such knowledge to anyone who had neither the authority nor the need to know. Further, they were instructed by the administrators of the organization to deny knowing anything about this information when subjected to questioning by the examiners. (Despite the realistic tenor of the above description, the subjects knew that this was an experiment and not a security investigation. Their participation in the research was completely voluntary.)

Whereas the Denial-of-Actual-Crime experiment aroused motives of personal security, the Denial-of-Information experiment depended upon the subjects' habits and training to withhold information which they believed was related to national security.

B. SUBJECTS

Twenty-six men, who had access to or worked with classified information, volunteered for the experiment. A minor selection bias may have operated, since no volunteer would have been accepted if his other work were to suffer as a result of losing time in this research. The age range was 20-35, intellectual capacity was well above average; cooperation was good. Motivation and involvement seemed adequate.

C. PROCEDURE

1. Test Questionnaires

Two questionnaires, each composed of 40 questions, were used with all subjects; two additional ones were used, after an interval of a week, in the retest of 11 subjects. The format of these questionnaires was the same as that described in the Denial-of-Actual-Crime experiment. Again, four critical questions were included in each form.

In the course of their regular duties, and within a day or two of the beginning of the research, the allegedly classified information was given to the group. As mentioned before, this information was completely fictitious. Several critical questions were formulated about this information. These were used in the questionnaires.

Additional critical questions were constructed around a new situation, one similar to that considered classified, and also fictitious. Questions concerning it served as controls, since subjects answering NO would be telling the truth. But in denying any knowledge about the allegedly classified data, the subjects would, in fact, be lying.

Two questions from each of the above situations were incorporated into each questionnaire.

As in the previous experiment, the task of the examiner was to determine which information was denied falsely and which answered truthfully. In contrast, however, the questionnaires used in this experiment were identical from subject to subject.

2. Test Sessions and Instructions

The subjects were first approached by the administrators of the organization, who, assuring them of complete anonymity, gave them some general information about the procedure and asked for volunteers. Subsequently, the subjects were referred to the supervisor of the experiment who gave them additional instructions before introducing them to the examiner. The directions and procedures of supervisor and examiner were not essentially different from those used in the Denial-of-Actual-Crime experiment. They are reproduced in Appendix C.

3. Test Environment

The polygraph was again transported to the organization which supplied the subjects for this experiment. The testing room was relatively large (approximately 15' x 15'), air conditioned, and quiet. All testing was done during the working day, from about 9 AM to 5 PM. Testing was completed within a two-week period.

4. Examiners

Two examiners, one of whom did the testing in the first week, were available for this experiment. The first examiner did not know any of the details of the fictitious situations about which the procedure was structured. His ratings and evaluations of the records were therefore uncontaminated by prior knowledge.

The supervisor, however, had to know the details of the test situation and was used for the examination of the subjects in the latter part of the experiment. It was requested that he also make an analysis, as objective as possible, of the records. In addition, the same raters who were used in the Denial-of-Actual-Crime experiment evaluated the records in this

experiment, without any knowledge about the experiment or about the questions that had been used. Their ratings were concerned only with the physiological reactions which were identified by number only.

D. RESULTS AND DISCUSSIONS

As in the previous experiment, the raters were required to evaluate the responses to four questions. Two of these questions were answered truthfully. Two received lying responses because the subjects would not admit the possession of classified information to persons who had no need to know. The rater had to identify the two lying reactions. In his choice of two reactions, the rater could get both correct, one correct, or none correct. The results for the two examiners and the two auxiliary raters are presented in Table 28, which also includes the related tests of significance. As Table 28 indicates, 11 of the 26 subjects were retested. In both testing periods two questionnaires, I and II, were used. The results of the two experienced raters (B and E) are grouped together for comparison with those of H and S, the lie detection trainees who were completely unaware of the procedures and questions used in this experiment. The Chi-squared values are listed when significance is attained (.05 level or better) and are designated by NS (not significant) when the appropriate level is not reached. In the retest, significance values were computed only for the totals, since the sample size was restricted and the expected frequencies were too small.

In general, the results are even more disappointing in this than in the Denial-of-Actual-Crime experiment. Other conclusions derived from this table parallel those obtained in the previous experiment. Thus, the more experienced raters seem to obtain slightly more accurate results. Again accuracy diminishes from the first to the second test session. The latter conclusion can be made both for the first and for the retest periods. Outside of these specific comments, the overall impression from the table is non-significance, or the inability of the raters to distinguish between those questions on which the subjects lied and those which they answered truthfully.

What, then, can be the explanation for these disappointing results? Could it be that these individuals made deliberate attempts to confuse the examiner by tampering with the physiological reactions being monitored by the polygraph? This was probably unlikely since, upon retest, the subsample reacted initially with about the same significance as it did in the first testing period. Was it due to boredom and adaptation? This is a possible explanation since the results of the first questionnaire (Test Session I) are generally significant while those of the second questionnaire fall to a chance level. This has also been the trend in the Denial-of-Actual-Crime experiment. This explanation would

TABLE 28

DISTRIBUTION OF CORRECT CHOICES FOR THE FOUR RATERS AND THE SIGNIFICANCE OF THE RESULTS IN BOTH TESTING PERIODS

TESTING PERIODS	TEST SESSION	RATER	Number Correct: Two	One	None	Chi-squared
FIRST (N=26)	I	B	10	15	1	8.89**
		E	5	19	2	NS
		Total	15	34	3	5.56*
		H	2	20	4	NS
		S	2	22	2	NS
		Total	4	42	6	NS
	II	B	6	17	3	NS
		E	3	21	2	NS
		Total	9	38	5	NS
		H	4	19	3	NS
		S	3	21	2	NS
		Total	7	40	5	NS
RETEST (N=11)	I	B	5	5	1	--
		E	4	7	0	--
		Total	9	12	1	9.32**
		H	5	5	1	--
		S	7	3	1	--
		Total	12	8	2	22.73**
	II	B	1	8	2	--
		E	3	6	2	--
		Total	4	14	4	NS
		H	2	8	1	--
		S	3	6	2	--
		Total	5	14	3	NS

** Significant at .01 level.
* Significant at .05 level.
NS Not Significant

emphasize the fact that the subject knew that he was in an experimental situation, and, despite the importance of the issues under interrogation, failed to become actively involved in the continuing test sessions.

One may even conjecture that the subjects became puzzled and suspicious about the "other" information, perhaps wondering if the

interrogation were a real investigation operating under the guise of an experiment. Though this was a possible explanation in the case of individuals who may have felt uncomfortable about some statements in their personnel applications, etc., the emphatic assertions of the administrators that this was an experiment and restricted only to volunteers, were sufficiently compelling to dispell any suspicions of a secret security investigation. A more reasonable variation of this explanation would stress the subject's interest in and possible anxiety over the "other" information he was being questioned about. Thus, he may have wondered whether this was a matter he forgot but should have remembered. Or, the "other" information may have set off a pronounced effort on the part of the individual to try to place it in his memory. Such effort would result in a physiological pattern different from that where no effort was exerted. But such an explanation would merely play up the fact that the original, supposedly classified information had lost its importance to the subject under continued interrogation.

One final possible explanation remains. In the search for a rapid diagnostic procedure, the testing had been limited to two questionnaires for this and for the Denial-of-Actual-Crime experiment. This may be too short for adequate lie detection purposes. One hears that all day interrogation sessions and repeated testing from one day to another are not uncommon practices among lie detector operators. By such standards, the 8-15 minute examination periods of these experiments are limited indeed. Furthermore, the examiners and raters in this experiment were not permitted to use a "No Decision" category. Whether they felt the evidence sufficient or not, they were requested to make a decision.

That these restrictions are sufficient to explain the low accuracy figures is doubtful. A diminution in accuracy was observed from Test Session I to Test Session II in this and in the Denial-of-Actual-Crime experiments. Consequently, additional testing of the same sort and on the same day would not have resulted in greater accuracy. Furthermore, in the limited retest phase of the present experiment, the results did not seem to indicate any dramatic increase in lie detection accuracy. Consequently, a repetition of the same type of procedure would not seem to be a sufficient remedy. A change to an altogether different testing procedure or principle might prove valuable. In fact, the lie detection field is ready for re-evaluation and overhauling. New ideas and new approaches are necessary if the initial promise of this potentially valuable technique is to be realized.

E. CONCLUSIONS

The accuracy of differentiating lies from truthful responses, in a group of individuals who denied possessing information they erroneously

believed classified, was low and disappointing despite the slightly better results obtained from experienced raters.

Of the several explanations considered, the one casting doubt on the adequacy of experimental procedures to test the validity of the lie detector would, in turn, ultimately lead to a loss of confidence in the claims of lie detection operators that their procedures have attained high accuracy in real life situations. A general re-evaluation of the traditional lie detection technique would seem necessary. Better yet, new approaches would more adequately revitalize the field and lead to new directions in examination and interrogation.

CHAPTER VI

COUNTERMEASURES EXPERIMENT

A. THE PROBLEM

The problem of countermeasures is related to the broader problem of survival, wherein living organisms are born with or develop tendencies to escape detection by an unfriendly agent. For example, we know that nature provides many of its creatures with camouflage or disguise which serve as protection against predators. Among humans the same tendency has an intellectual counterpart, ---deception.

Any procedure aimed at breaking down the deception barrier will inevitably arouse a counter tendency aimed, in turn, at the new procedure. The countermeasure is, strangely enough, the same old weapon, ---deception. The attempt now is to deceive the agent probing deception.

In all training manuals on deception detection, the operator is warned to be on the alert for guilty suspects who make attempts to "beat the lie detector." The usual attempts are manipulation of the instrumental attachments or tension in the musculature, both of which produce artifacts in the records. The untutored criminal has neither the theoretical background nor sufficient time to develop scientific countermeasures to help him escape his embarrassing predicament. When, however, stakes are high, as they well might be in intelligence and counterintelligence operations, one nation could train its operators in lie detection countermeasures if it knows that lie detectors are used by its enemy. Similarly, an intelligent criminal with intelligent advisors may present a challenging problem even to an experienced lie detector operator.

This discussion and the experiment to be described are based on the hypothesis that there is or can be developed an effective countermeasure to every human attempt that threatens the survival of other human beings.

B. COUNTERMEASURES IN LIE DETECTION

Crude muscle movements, belabored breathing, and manipulation of the instrumental attachments are the usual artifacts for which the operator is alerted by his training. A more sophisticated procedure would involve the use of drugs. Conditioning procedures might be attempted. A pilot experiment along these lines (Block et al., 1952) did not prove particularly successful, but improved methodology and a more imaginative approach would undoubtedly produce more definitive results.

In addition to those suggested above, there are several possible countermeasures, each depending upon the voluntary control of the person tested. In one sense these are desirable methods, since the person can change tactics at will, should the occasion require quick reaction to an altered test situation. (The use of nonconscious countermeasure procedures is not the concern of this investigation.)

The three countermeasures selected for study are: 1. Modified Yoga, 2. Muscle Tension, and 3. Exciting Imagery. The names are almost self explanatory. In the Yoga procedure, the emphasis was on the separation of the self from outside stimuli and the maintenance of an abstract frame of mind that would still enable the subject to respond to questions mechanically, but without affect. Intense concentration upon a specific object was not encouraged in this procedure since that was to be stressed in the third method. Elimination of affective reaction was the prime consideration.

The Muscle Tension procedure was selected as particularly apt for those subjects who could not control their affective reaction to carefully structured questions. The attempt to control musculature would serve to drain and channel the emotional reaction into a natural expenditure of energy. It should be clear that selection of a muscle site should be made with due attention to its capacity for movement without the examiner's detection. For this experiment the subject was instructed and trained to press his toes against the soles of his shoes which were resting flat upon the floor.

With proper instrumental controls, tension in the leg muscles can be detected. There are, however, muscles which ordinary instrumental controls cannot monitor so readily. It was felt that training the subject in "toe control" was easier and less time consuming.

Muscle control as a countermeasure serves two diversionary functions. It forces the subject's attention to the muscle control task and thus lessens the impact produced by the emotional response of lying. Moreover, the effort expended in fulfilling the muscular task produces a relatively natural physiological response whenever required.

Of the three countermeasures, the Exciting Imagery technique seemed to hold the most promise. In the first place, it is controllable by the subject. Secondly, it is not detectable; no instrumental devices have been developed that can monitor this voluntary response. Simply stated, the subject was to reproduce, in his mind, an exciting image or situation at the appropriate places in the interrogation procedure. The exciting image was to be one which the subject knew could get him excited or upset.

Whatever the countermeasures used, it should be obvious that the scientifically minded lie detector operator would try to develop measures either to detect countermeasure activity on the part of the subject or to neutralize such efforts. In line with such thinking, a search would be made to find a means of monitoring a physiological system that would most likely be triggered by the effort the subject exerted in his countermeasure activity. Where such a physiological monitoring system could not be discovered or readily attached, statistical controls might be utilized to distinguish between deliberate countermeasure attempts and random activity that occasionally appears to be deliberate.

C. PROCEDURE

1. Subjects

Twenty students were selected from among those who served as subjects in the Simulated Theft experiment. Because the directions for the countermeasure procedures were somewhat involved, it was decided to use subjects who had more experience in a polygraph experiment, since it would take them less time to master the evasion techniques. Furthermore, this experiment provided the final test for the cash prizes promised the subjects during the recruitment stage. The subjects selected were the finalists from among whom nine were to win cash prizes.

Of the 336 subjects used in the Simulated Theft experiment eight were selected because they were easiest to classify correctly, eight because they were the most difficult, and eight because they were "midway" between the two extreme groups. Half of these reported that they experienced more than average anxiety during the various phases of the experiment. The other half experienced very little tension or anxiety, ---the so-called "cold blooded" type. Of those contacted, twenty were available within the time limits set by the test schedule, and these were used in the experiment.

2. Examiners

Two examiners were involved in this phase of the research. Each tested 10 subjects.

All testing was done in the one-way vision examining room.

3. Instrumentation

The polygraph, as described in Chapter II, was used.

4. The Numbers Test

In the Numbers Test the subject chooses a number which the examiner tries to discover despite the subject's denials. This test was chosen because it is easy to control and to administer. As used in this experiment, the subject picked a folded slip of paper from among many in an envelope. He noted the number on the slip which he put into his pocket. He was requested to say NO to all questions of the type "Was the number ____?" even though the number on the slip may have been asked. Specially prepared number lists were used by the examiner, who did not know the number chosen by the subject.

Since there were four test conditions, namely, Control, Yoga, Muscle Tension, and Exciting Imagery, different lists were needed for each situation. One list included the numbers 1 to 10 and this was used for the first or control session. The other lists, in order, were 10-19, 20-29, and 30-39; they were used with Yoga, Muscle Tension, and Exciting Imagery conditions respectively. In each list, the numbers were repeated three times. The experiment was so controlled that the supervisor of the experiment knew what number the subject selected but the examiner did not. Furthermore, the specially constructed number list contained the critical number (i. e. the chosen one) at appropriate intervals to permit better evaluation of the responses. Five different forms of the lists were used, each prepared for a different critical number. There were, then, twenty number lists, five for each test condition.

After each test condition was completed, the subject put the slip into an envelope which he kept in his possession during the experiment. Since the Muscle Tension and Exciting Imagery situations involved the use of decoy numbers, the subjects indicated on a slip of paper the particular number on which they chose to use Muscle Tension or Exciting Imagery. In other words, the subjects were to simulate lying by inducing Muscle Tension on a number different from the one they had chosen, thus hoping to "fool the examiner" and to have him decide incorrectly on the decoy number instead of the number on the slip. All these slips were put into the envelope by the subject who signed, sealed, and addressed it to the director of the project.

5. The Test Situations

The extensive directions for each of the four test situations are in Appendix D. Only a brief resume will be included in this section.

In all four test situations, a numbered slip was chosen by the subject who denied to the examiner that he had picked that number. In the

first or Control session, no special conditions were involved. This was to represent a basal reaction level against which the results of the next three sessions were to be compared.

The Yoga instructions were given in an attempt to help the subject respond without affect or emotion to the stimuli presented to him. He was to assume an impersonal, "far away" attitude to the questioning, denying the number without personal involvement or care.

The Muscle Tension instructions taught the subject to tense his toe muscles on a number different from that he had chosen. In this way, it was anticipated that the effort and excitement involved in tensing on the appropriate decoy number would simulate an emotional involvement with that number and produce a reaction that could be mistaken for a lie response.

The Exciting Imagery condition, by requiring the subject to bring to mind a situation which, from experience, he knew could excite him, was intended to help him produce an emotional storm at will. As in the Muscle Tension experiment, the subject was instructed to attach this exciting thought or image to a number different from that he had chosen. Again the purpose was to simulate an emotional reaction to mislead the examiner.

In fact, the special test conditions can be interpreted in terms of camouflage or "jamming." The Yoga procedure stresses non-involvement or no reaction so as to create the impression of innocence. The other two procedures intend to confuse the examiner by "jamming" his graph with irrelevant reactions in areas not related to the critical issue. The effort expended by the subject is expected either to produce generalized "emotional" reactions that would make any interpretation difficult, or to lead the attention of the examiner towards areas (or issues) which cannot seriously incriminate the subject.

To help guarantee that these evasive or countermeasure techniques were used properly, a short training session was provided for before the tests. The examiner began the test proper only after he was satisfied that the subject was able to control his physiological reactions.

One final point bears mentioning. The nature of the test situations was such that the order of testing was determined. The Control test could have only come first, for it is difficult to see how a subject could give a "normal" or "natural" response so soon after learning how to control his reactions. It is also clear that the Exciting Imagery test had to come last. If used earlier, there would have been a definite danger that these exciting thoughts would spread to later sessions. The

Yoga procedure could not have followed either the Muscle Tension or the Exciting Imagery tests since it would have been difficult to relax physically and mentally after being trained in such novel techniques with which to "beat the lie detector." It is fairly clear that the order of testing in this experiment was a natural one in that it minimized possible confounding effects among the test conditions.

D. EVALUATION OF RECORDS

The records were evaluated by two raters who neither participated in the experiment as examiners nor knew its design. Consequently, their ratings were not contaminated by prior knowledge, which, it was felt, would have changed their criteria of interpretation. After a considerable lapse of time, one of the examiners in the experiment also rated the records.

The basic set of instructions given to the raters was: " Pick out the number on which the person is lying." So that the efficacy of each of the physiological indices could be determined, the ratings were made on one index at a time. In other words, the respiration index was evaluated first for all subjects, each rater attempting to select the incriminating number from the respiratory patterns alone. Then, the plethysmograph records, and finally, the psychogalvanic responses were evaluated. As would have been done in a real lie detection procedure, the final task was the evaluation of the records as a whole. From these ratings, four accuracy scores were produced--for respiration, finger volume, psychogalvanic response, and for "total," the latter being obtained from an evaluation of the total record with all three physiological responses available for examination. The first three of these scores were relatively free from rating interaction effects, since separate recording forms were used for rating the different indices.

Once these ratings were completed for all four test situations, additional instructions were given for the evaluation of the Muscle Tension and Exciting Imagery data. It was in these two situations that the subject was taught to use "decoy" numbers in order to lead the examiner to select as incriminating a number other than the one on the slip. A slip of paper with three numbers on it was attached to each test record. One of the numbers was the incriminating number; the others were the two from which the subject selected the decoy number on which he either tensed or elicited imagery. Added were the instructions:

> "Of the three numbers clipped to each record (for each subject and for each record) pick out the
>
> (a) lie
> (b) simulated lie
> (c) neutral number."

The three numbers were printed in a random order on the attached slips of paper. Again, the respiration records alone were evaluated for all subjects before such ratings were done for the other physiological responses. And finally, the "total" records, with all indices available for examination, were rated under these instructions.

The aim of this rating procedure was to discover how differentiable the real lie would be from the simulated lie once the rater was alerted to the presence of camouflage, or, the use of decoy numbers.

E. RESULTS

1. Detection of the Incriminating Number (Lie)

The first set of results report the relative accuracy with which the raters picked out the number on which the subjects lied. Since the examiner's ratings did not differ from those of the two independent raters, the accuracy scores of all three were averaged and rounded off to a whole number. The accuracy scores under each of the four test conditions are presented in Table 29. The results for the separate indices--respiratory, plethysmographic and the psychogalvanic--are

TABLE 29
ACCURACY SCORES IN DETECTING LIES (NUMBERS TEST) FOR THE THREE PHYSIOLOGICAL INDICES UNDER FOUR CONDITIONS OF TESTING

TEST CONDITION	PHYSIOLOGICAL INDICES			
	Resp	Pleth	PGR	ALL
Control Session	7	8	16	15
Yoga (Modified)	5	8	15	16
Muscle Tension	2	2	4	2
Exciting Imagery	4	3	3	5

reported, along with the accuracy scores when all indices were available for evaluation. The accuracy score is the number of subjects whose lies were correctly identified. Thus, in the control session an average of seven of 20 subjects were detected in their lies by means of the respiratory response alone; eight were detected through their plethysmographic

records; 16 through the psychogalvanic response; and 15 were detected when the entire set of physiological responses were available to the rater for evaluation. As in the Simulated Theft experiment, greatest accuracy was observed for the psychogalvanic response.

The data of Table 29 indicate that the Yoga procedure was not effective as a countermeasure technique. The accuracy scores for this condition did not differ from those of the control session. Thus, an attempt to be calm, to have one's mind abstracted, to disregard the emotional impact of the questioning procedure, as was practiced by the subjects under the Yoga instructions, was not reflected by a diminution in the physiological responses to the critical stimuli. It is possible that the Yoga technique, to be effective as a countermeasure procedure, demands more practice and training than was available in the short pre-test period.

Quite different results were obtained with the Muscle Tension and the Exciting Imagery test conditions. In both there was a marked drop in accuracy for all physiological indices, considered individually or combined. From the uniformity and consistency of the scores, it seems that both of these countermeasure procedures were about equally effective. The reduced accuracy ranged from 1/2 to 1/5 that reported for the control session. For example, the accuracy scores for the Muscle Tension condition were about 1/4 as large as those for the control session for each of the three physiological indices.

Since the sample was small no extensive testing for significance was considered. However, the results for PGR and for "All Indices" showed significant differences in accuracy (.01 level) between the control session and both the Muscle Tension and Exciting Imagery test conditions. Furthermore, the effects of these two countermeasures can be observed in the non-significant accuracy scores they produced. Since the chance probability of detecting a lie is .10 (there are 10 numbers to choose from), the cumulated binomial probability of detecting seven lies or more in 20 individuals is about .002. Thus, all accuracy figures for the control session may be considered highly significant. In contrast, four correct critical number selections represented the highest accuracy score for the three indices in either the Muscle Tension or the Exciting Imagery test condition. The probability value associated with four or more is .13, a non-significant value if the ordinary significance levels are used as the criterion. Thus, the countermeasure technique lowered the accuracy scores from a highly significant level to a level which cannot be considered as indicating better than chance results.

The fundamental conclusions of this section are: (a) Muscle Tension and Exciting Imagery countermeasures were effective in deceiving the lie detector operator; (b) the Yoga technique did not function as an effective countermeasure under the conditions of this experiment.

2. Effectiveness of Decoy Numbers

It must still be established that the impressive results reported in the previous section have, as their effective cause, the countermeasure procedures themselves. To accomplish this, the frequency with which the decoy numbers were mistaken for the real numbers must be analyzed. This is done in Table 30, where the data for the control sessions have been included, together with the accuracy figures from

TABLE 30
FREQUENCIES OF DETECTING REAL NUMBERS CORRECTLY AND DECOY NUMBERS INCORRECTLY IN THE MUSCLE TENSION AND EXCITING IMAGERY CONDITIONS

	MUSCLE TENSION				EXCITING IMAGERY			
NUMBERS	Resp	Pleth	PGR	All	Resp	Pleth	PGR	All
Decoy	6	7	12	14	3	8	12	11
Real	2	2	4	2	4	3	3	5
SUM	8	9	16	16	7	11	15	16
CONTROL	7	8	10	15	7	8	16	15

Table 29. The first line of Table 30 gives the frequency with which the decoy numbers were mistaken by the raters for the real numbers. The second line indicates the frequency with which the real numbers were correctly identified under the two countermeasure procedures. The sum of these gives a set of frequencies which are remarkably similar to the accuracy figures for the control sessions which are listed in the last line of the table.

Table 30 indicates that the decoy numbers used in the two countermeasure situations were clearly effective since they appeared as lies to the rater. In seven of the eight comparisons (lines 1 and 2), the decoy numbers were designated as lies more frequently than the real numbers. If the first two lines are summed, then the decoy number was considered a lie 73 times in comparison to the 25 times the lie itself was detected, a ratio of three to one. Thus, the two countermeasure procedures were not merely effective in decreasing the accuracy of the raters, but more important, they were able to direct the raters' decisions in a specified direction, away from the real lie.

3. Accuracy of Raters under Knowledge Instructions

But how would the raters perform if they were given knowledge about the specific test instructions? More important, how accurate would they be if they were given three numbers and told to detect the real lie, the simulated lie (decoy number), and the neutral number? Table 31 gives the accuracy of the raters in detecting the real, decoy, and neutral numbers for all these physiological indices.

TABLE 31
ACCURACY SCORES FOR DETECTING THE REAL, THE DECOY, AND THE NEUTRAL NUMBERS IN THE MUSCLE TENSION AND EXCITING IMAGERY CONDITIONS

PHYSIOLOGICAL INDEX	NUMBERS	MUSCLE TENSION	EXCITING IMAGERY
Respiratory	Real	9	6
	Decoy	8	6
	Neutral	7	8
Plethysmographic	Real	8	8
	Decoy	7	6
	Neutral	10	9
Psychogalvanic	Real	7	9
	Decoy	7	8
	Neutral	16	15
All Indices Available	Real	9	10
	Decoy	8	9
	Neutral	15	14

The first impression gathered from analyzing the table is that the accuracy of identifying the real number is practically the same as that for identifying the decoy. Thus, for the respiratory index, the real number is identified correctly in nine out of 20 instances under Muscle Tension conditions, and the decoy in eight out of 20. Similar comparisons reveal the same result for the different physiological indices and for the two countermeasure situations. With chance expectation of about seven

(20/3), the obtained accuracy figures, averaging slightly above seven for the three physiological indices, are hardly impressive. When all indices were utilized in the evaluation, the accuracy score for the real and decoy numbers increased to about nine, a very limited improvement.

The validity of the above results is corroborated by the accuracy scores for the neutral numbers in the analysis of the psychogalvanic responses. These scores are 16 and 15 for Muscle Tension and Exciting Imagery respectively, scores almost identical with those obtained under control conditions. This is to be expected, since the differentiation of the neutral numbers from both critical and decoy numbers should be as clear as the differentiation between the critical and other numbers in the control session. These results would seem to preclude any explanation that would cast doubt on the adequacy of the obtained data.

It is difficult to evaluate what underlies this rating procedure. Once the rater is told that the subject is trying to deceive him, it is likely that he will change his rating criteria. It is even possible that he would consider the larger or the more intense of two reactions as due to countermeasure influence and rate the smaller or "weaker" reaction as the lie, the reverse of his usual rating procedure. The rater also might try to estimate how much effect, for example, Muscle Tension would have on the physiological response. Unless such changes were based upon established physiological principles clearly differentiating lies from simulated lies, his decisions would assume a chance character. And this is what seems to have happened where approximately equal accuracy scores for the real and decoy numbers were obtained, with neither of these scores attaining statistical significance.

F. DISCUSSION AND CONCLUSIONS

It is reasonable to assume that the guilty criminal under a lie detection examination would attempt to "beat" the lie detector if he knew the proper procedure. Many would try even without such knowledge. All such attempts tend to make the job of the lie detector operator more difficult, since he must decide whether irregularities in the records are deliberately introduced by the guilty suspect or unconsciously produced by an innocent, and possibly anxious, individual.

Three generic countermeasure procedures were evaluated in this pilot experiment. One stressed control over the emotions with an attempt to assume an impersonal and distant attitude insofar as the questioning was concerned. Another emphasized the value of diversion by doing something concrete and peripheral, thus channeling energy into a natural outlet. The third also was a diversionary technique, aiming to keep the matter under questioning secondary to another possessing high emotional value.

The special advantage possessed by the latter two techniques was that of producing physiological reactions when and where desired by the subject. Also serving to diminish reactivity to the critical questions was the fact that the subject had a specific task to perform. This also occupied his mind during the questioning.

Of the three lie detection countermeasures the Yoga procedure was least successful and could not be differentiated from the results of the control test. Abstraction and separation of the emotional associations from thoughts and memories was not readily achieved in the short training session available to the subjects. There is considerable doubt that this procedure would prove effective even with much greater training. If it would tend to leave the mind of the suspect relatively free from agitation and emotional content, then stimuli differing in emotional value would be more clearly differentiable in their physiological effects than if the mind of the suspect were in a turmoil during the entire questioning period. Possibly only the experts in this method of concentration could respond intelligently to questioning without in any way experiencing an emotional involvement in the issues under investigation.

The Muscle Tension and Exciting Imagery countermeasures proved equally effective in deceiving the interrogator. The use of these procedures reduced the accuracy of detection from high significance to chance levels. The control exhibited by the subjects over the test situations, even with the limited training they received, showed that these methods might be particularly valuable countermeasures in the hands of an intelligent and resourceful person. He could, in effect, direct the examiner's attention where he wished, thus giving himself control over the test situation.

The most important result underlying the successful use of the countermeasures is the inability of the lie detection operator to distinguish the physiological reactions in deception from those in simulated deception. In other words, the tensing of the muscles and the bringing to mind of exciting thoughts produce a physiological pattern that cannot be distinguished readily from that found in lying. It could be that refined methods of analysis may ultimately bring out important differences between these two types of reaction. It is also possible that these difficulties will be bypassed with the development of new lie detection procedures. Or, techniques which are relatively insensitive to such simple countermeasure procedures may be discovered. But until such procedures are developed, or new indices discovered, the use of ordinary lie detection procedures will be of questionable value in the examination of intelligent subjects trained in effective countermeasures.

One objection still remains. In the Numbers Test, no vitally important personal issues are at stake. When the subject is instructed to lie about a number, it is not a real lie but a simulated lie. Thus, no real lying is involved in these experiments.

This is, indeed, a valid objection and one that cannot be dismissed lightly. The only answer that can be given that these experimental situations generate physiological response patterns that seem indistinguishable from those found in the lying responses of guilty individuals being interrogated about their participation in a specific crime (Inbau and Reid, 1953). Apparently, the mere denial of a true state of affairs generates similar physiological responses whether they be in experimental or real life situations.

The conclusions, then, are clear and decisive. Countermeasures can be developed which deceive and confuse the lie detection operator in his evaluation of the records. Even when the rater knows that procedures are being used and even though he is given a fair opportunity to distinguish between the lie and simulated lie, his accuracy is seriously impaired. If these experimental situations can be validly extrapolated to interrogation conditions in real life, the value of conventional lie detection procedures would be seriously jeopardized.

CHAPTER VII

DISCUSSION, CONCLUSIONS, AND RECOMMENDATIONS

The series of experiments described in the earlier chapters provide some answers to a wide variety of problems underlying the theory and practice of lie detection. Fundamental among these are the problems of validity and objectivity, issues the resolution of which was necessary before computerizing the lie detection data could be attempted. For, despite the faith of many scientists in the capability of the modern computer to handle many difficult and pressing problems, the primary requirement in lie detection is that the output of the lie detector (and, consequently, the input to the computer) be a valid discriminator between truth and deception. The point at issue was not the ability of the lie detector operator, but the lie detection record itself, which had to possess the telltale marks of guilt or innocence which were clearly manifest, even to those who had not examined the suspect. Objectivity, in other words, had to characterize the records.

As a problem, the feasibility of computerizing the lie detector placed a number of restrictions on these experiments. In the first place, the outputs of the polygraph components had to be in the form of electrical signals. Consequently, to record changes in the respiratory and cardiovascular systems, transducers were necessary. Although these appeared to be adequate substitutes for the traditional pneumatic systems used to record respiration and blood pressure, the recordings in these experiments cannot be considered to be completely isomorphic with those produced by the commercial polygraph. Secondly, in order to guarantee objectivity, interaction between the lie detector operator and the person tested had to be minimal. Therefore, there was no preliminary interview between the operator and the subject, as is the practice in actual lie detection investigations. Neither was there, between the lie detection sessions in this experiment, any provision for a period of information seeking or giving for purposes of clarification. On both counts, then, the procedures in this research did not parallel actual lie detection practice and, consequently, do not invalidate the claims of lie detection experts who often express great confidence in the accuracy of their instruments.

Scientific and practical considerations limited the nature and extent of the interrogation sessions in these experiments. Furthermore, the procedure was kept constant from subject to subject. In actual lie detection investigations the questioning procedures are tailor-made to fit the person being interrogated. They are also very flexible and are altered to meet any new changes and developments taking place as a result of earlier testing. However attractive such flexibility may be, there is the critical disadvantage in not being able to disentangle the contributions of

machine and expert in the lie detection process. For computerization, it is precisely the contribution of the machine that had to be evaluated.

A. VALIDITY CONSIDERATIONS

With these restrictions in mind, one can turn to the problem of validity as it was faced in this series of experiments. Fundamentally, the issue involved the accuracy with which the polygraph records could be used to discriminate either (a) between people who lied and people who told the truth, or (b) between lying and truthful responses in the same person. Though these appear to be similar, they are, in fact, different problems.

Closely related were the critical issues of realism and countermeasures. It is true that a fully controlled lie detection experiment could never guarantee full realism, since ethical considerations would not permit the commission of a crime for the sole purpose of studying the behavior of the participants. Such an experiment, nevertheless, should attempt to attain as much realism in its design as is possible within the bounds of morality and propriety. In this way, the validity of the lie detection experiment could not be attacked on the grounds of triviality and obvious non-applicability to a real life situation.

As for the issue of countermeasures, a systematic study of invalidating factors is a necessary component of any validity study. In lie detection, one of the most serious and challenging sources of invalidity is the deliberate attempt on the part of the guilty person to mislead the operator by any means at his disposal. A clarification of the extent to which this is possible was an essential feature of this study.

In all, four experiments were conducted, three on several aspects of realism and one on countermeasures. The most extensive among these was the Simulated Theft experiment in which the lie detector examiner had to identify the role of each of three persons involved in a simulated theft. These were the Thief, the Lookout, and the Innocent Suspect. All five examiners used in this experiment were able to differentiate among these roles with highly significant accuracy. When these same examiners evaluated the records obtained by their colleagues, their accuracy was equally high and significant despite the fact that they worked from the records alone. These are important conclusions in that they emphasize the existence of valid indices of deception in the graphic records themselves.

The records contained three sources of information for the lie detection examiner (or independent rater), namely, the respiratory,

plethysmographic, and psychogalvanic response patterns. Since all three were rated, an analysis of the results revealed that the psychogalvanic response attained greatest accuracy in discriminating among the Thief, the Lookout, and the Innocent Suspect. Respiration yielded the lowest accuracy. To some extent these results were due to the greater complexity of the respiratory and plethysmographic response patterns, which made it more difficult for the raters to evaluate them with objectivity. Further, the simulated situation, in contrast to a real crime, undoubtedly aroused less emotional involvement. Consequently, it is reasonable to assume that the psychogalvanic response would lose less of its discriminating power than the respiratory and plethysmographic reactions, these being somewhat less sensitive indicators. Inbau and Reid (1953, p. 102) seem to come to the same conclusions but they emphasize that the situation is reversed in actual criminal cases, where they contend that the blood pressure-pulse and respiration recordings are more dependable than the psychogalvanic reactions.

Two additional experiments were conducted to study the problem of realism, the Denial-of-Actual-Crime and the Denial-of-Classified-Information. In the first, patients on parole were instructed by their therapists to deny any participation in crimes for which they had been originally apprehended. It was assumed that these patients, while in the process of rehabilitation, would naturally refrain from disclosing such socially prejudicial information, especially to unauthorized individuals. Their physiological responses, while denying the actual crime for which they were arrested, were compared with the responses generated by denying participation in a crime which they had not committed. Under these test conditions, the lie detection operators and the chart analysts attained a much lower degree of accuracy than that achieved in the Simulated Theft experiment. A possible explanation for these disappointing results could have involved the comparison stimulus, that is, the "other" crime which the parolees had not committed. Since they had already confessed to the crime for which they were apprehended and since they had discussed that situation in therapy, it would seem that the fear of detection, or feelings of shame and guilt, would have diminished appreciably by the time they had been questioned in this experiment. In contrast, the "other" crime may have aroused associations involving their friends' activities, and consequently stirred up anxieties as to whether the test procedure was an experiment or an actual investigation instigated by the law authorities. In other words, the comparison stimuli may have aroused greater emotional involvement and anxiety than the questioning about their former transgressions. Implications in new crimes were, of course, a greater threat than discussions about old crimes. In effect then, the test situation could have become too realistic for these inherently suspicious subjects who may have misinterpreted the

purpose of what was planned as an exciting but innocuous experiment.

The above explanation for the inadequate results obtained from the Denial-of-Actual-Crime experiment is one that assumes the validity of the lie detection procedure. The other alternative would be to consider the results as questioning the value of these traditional interrogation aids. The issues involved in such opposing explanations were reemphasized in the results of the second experiment which was hopefully designed to further test and evaluate the factor of realism in lie detection research. In this experiment, the Denial-of-Actual-Classified-Information, workers in an organization that handled confidential material were instructed to deny knowledge of any classified information that had been given to them. Actually, they were given fictitious material which they assumed to be classified information. As in the Denial-of-Actual-Crime experiment, the physiological reactions to this allegedly classified material were compared with the reactions to other fictitious material which, however, was not described as classified.

The results of this experiment were even more disappointing than those of the Denial-of-Actual-Crime. And again the explanation was similar in nature. It was relatively easy for these subjects to deny knowledge of the allegedly classified information to interrogators who had no need to know. But when questioned about new material, these subjects may have been puzzled and then made serious efforts to search their memories for information which they "feared" they had forgotten. It may have been more serious to them to have forgotten something important than to routinely deny a fact they were instructed to withhold from those who had no need to know such information. As a matter of fact, the recruitment instructions given to these subjects were not under the control of the present investigator. The management insisted that the subjects be instructed that their participation was completely voluntary and anonymous and that the investigation was completely experimental in character. Whatever may have been the frame of mind of these subjects, it was unlikely that they considered the test as a threat to their security on the job.

But these, again, are the exonorating explanations, excuses for the inadequate performance of a technique assumed to be valid. In one sense, at least, these conjectures get some support from the Countermeasures experiment in which the effects of deliberate attempts to distort the records were evaluated. It was found that induced muscle tensions and exciting associations could produce physiological reactions which examiners and raters mistook for lies. It is possible then, that the "other" crimes about which the patient was questioned in the Denial-of-Actual-Crime experiment aroused exciting or uncomfortable associations which confused the interpreter of the records.

In like manner, interrogation about the fictitious information in the Denial-of-Classified-Information experiment may have aroused a puzzled attitude and troubling associations which, in turn, may have produced disturbed physiological responses. Regardless of the adequacy of this explanation, the success of the countermeasures in deceiving the lie detector operator can only point to the vulnerability of the lie detection procedure. Should these countermeasure procedures prove equally effective in actual criminal interrogations, the value of the traditional lie detector, in the examination of intelligent and knowledgeable criminal suspects, would be seriously questioned.

B. OBJECTIVITY CONSIDERATIONS

Independent raters, working only with the records, obtained the same degree of accuracy as the examiners who tested the subjects in the Simulated Theft experiment. This would seem to indicate that the test procedure produced relatively objective records. However, when the discriminant functions were calculated for the examiner and for the raters, all of whom evaluated the same set of records, these mathematical expressions were found to differ appreciably in their assignment of weights to the three physiological reactions. Since the discriminant function was developed from the individual discriminant scores which were computed for each physiological system, the objectivity of these scores was determined by correlation procedures. Thus, in a set of data which were rated, for example, by an examiner and a rater, there were two respiratory discriminant scores (one for the examiner, the other for the rater), two plethysmographic discriminant scores, and two psychogalvanic discriminant scores for each subject. The correlation coefficients between these pairs of discriminant scores provided an estimate of the objectivity of the ratings. In general, only the psychogalvanic correlation coefficients were sufficiently large to provide confidence that the recordings were rated with an adequate degree of objectivity for this type of analysis (visual estimates). The coefficients for the respiratory and plethysmographic discriminant scores were low and indicated that these physiological patterns are not rated with consistency. Undoubtedly this is due to the complexity of these patterns and, in part, to the relatively moderate disturbance produced within them by the experimental situation.

Two antithetical conclusions necessarily follow. Either these two systems are not good indicators of lying in experiments such as these, or the visual analysis of the records is inadequate to identify the subtle variations in the physiological patterns. The second alternative presents no inherent difficulties for the future use of these components in a lie detection system. The solution would be to eliminate all

visual analyses of the records and substitute objective measurement for such judgments. The ultimate goal would be automatic electronic analysis of the electrical outputs of the lie detector.

Suppose, however, that the plethysmographic and respiratory systems are not valid lie detectors in experiments such as those completed in this research. A number of reasons may be considered why this is so. Perhaps the emotional context about which the subject is questioned has to be a real crime before these two systems can be used for lie detection purposes. If so, then ordinary experimental verification of the validity of these systems could never be determined, since real crime situations, under the conditions of which these instruments would have to be evaluated, are the most difficult to control experimentally. And yet the ratings of the subjects and the observations of the supervisors in the Simulated Theft experiments indicated that there was appreciable emotional disturbance produced within the subjects by the "theft" and test situations. In fact, early experimentation with respiratory and blood pressure seemed to indicate that these physiological systems did not require massive emotional upheavals to produce differential responses to deception and truth.

It is still possible that the transducers used for recording the changes in respiration and blood volume of the finger may not have yielded the same types of records as those produced by the pneumatic systems usually found in the commercial polygraph. This may be true since, for example, the plethysmograph was substituted for the more conventional blood-pressure cuff because the latter could not be used for more extended periods of interrogation. However, recent investigators in the field have considered the plethysmograph an adequate substitute. Furthermore, if computerization was to be the ultimate goal, the output of the lie detector had to be in the form of electrical signals. Nevertheless, since there are many different types of transducers to measure respiratory and cardiovascular changes, further experimentation is desirable in order to determine the most accurate and dependable sensing and recording systems that can be developed.

Finally, the low objectivity in rating the respiratory and plethysmographic reactions may be due to the absence of a specific, well defined criterion. The concept of "general pattern changes" suggested by Joseph (1957) may have lacked sufficient precision, despite the fact that it is the criterion most widely used in actual interrogations, especially when the "classic" patterns fail to appear. In general, a vague criterion lowers objectivity as it might well have done in the evaluation of the test records obtained in this research. It becomes necessary, then, to search for additional and

more specific criteria of disturbance in these physiological records. Since the changes in these complex patterns are difficult to detect by visual analysis alone, instrumental (and therefore objective) procedures need to be developed to help in the analysis of the wave pattern changes and in the measurement of specific characteristics of these physiological responses.

C. LIE DETECTION AND THE COMPUTER

Under one aspect, the application of the computer to lie detection seems premature in view of the many problems associated with the detection of the "lie response" and the many others related to its measurement. The problems created by the results of the counter-measures experiment would seem to be surrounded with urgency aspects and therefore demanding of higher priority insofar as research is concerned. It would be patently unwise to develop a computer facility for lie detectors only to discover that its results could be controlled and manipulated by an intelligent criminal.

And yet it appears that some type of computer system will be necessary in lie detection. In the beginning stages of application, it may not be possible to use the computer to develop objective decisions of guilt or innocence. It has been pointed out that there is insufficient homogeneity among the various discriminant functions obtained by different raters of the same set of records. Without stable functions of this type, an objective and valid Yes-No, Guilt-Innocent decision is impossible. Perhaps the precise and objective differential measurement of the physiological patterns involved in truth telling and in deception is the most valuable contribution a computer system can make to the lie detection field. This, it has been pointed out, cannot be done with objectivity by lie detector operators in their ordinary visual examination of the records.

D. CONCLUSIONS FROM THIS RESEARCH

Despite the diverse character of the experiments completed for this research, a few general conclusions seem to bring order and meaning to the variety and specificity of the individual results.

In the first place, there seems to be sufficient validity (and even objectivity) in the Simulated Theft experiment to warrant confidence in the lie detection procedure as a possible aid to the interrogation process. All the examiners, trained as lie detection operators, were found to attain significant accuracy in identifying the Thief, the Lookout, and the Innocent Suspect in a Simulated Theft experiment. Raters, working with only the physiological records, attained the same degree of accuracy as the examiners who tested the subjects. The Questionnaire

procedure and the modified Peak-of-Tension test were most effective; the proper use of the Association test also showed significant discriminating power.

The second conclusion is concerned with the witness or suspect who is being interrogated. If his mental attitudes and associational processes can be controlled by the operator, especially within the brief period of the lie detector test, improvement in lie detection accuracy will most likely result. This is a reasonable hypothesis to explain the differences in the results with the Simulated Theft, the Denial-of-Actual-Crime, and the Denial-of-Classified-Information experiments. It was not possible to control the mental attitudes of the subjects in the latter two experiments. Furthermore, the Countermeasures experiment indicated that the greater the control the subject could exert over his own attitudes and responses, the greater contamination he could produce in his records and thus elude detection. The modified Peak-of-Tension test, in contrast to the usual Free Association test, further illustrates that greater accuracy is achieved when the examiner can control the character of the suspect's associational processes.

Thirdly, the intelligent suspect is capable of controlling his reactions to such an extent that he can elude detection in the ordinary lie detection procedure, provided that he has been trained in the use of effective countermeasures.

In the fourth place, the various physiological indices differ in lie detection accuracy. One important reason for such differences in accuracy is that the indices do not possess the same degree of objectivity. In this research, the psychogalvanic response was found to be most effective; it was also rated with the greatest degree of objectivity. But objectivity was low for the respiratory and plethysmographic reactions.

Finally, in view of the last two conclusions, this research did not find sufficient evidence to encourage the development of a computer program for immediate, on-the-spot decisions as to the guilt or innocence of a suspect tested on the lie detector. However, in view of the low objectivity in rating respiratory and plethysmographic responses, the use of computer techniques to differentiate among the complex physiological patterns obtained in lie detector examinations, and to develop new indices or criteria of deception, is indicated.

E. RECOMMENDATIONS

1. In view of the ease with which subjects can learn countermeasure techniques and thus successfully elude detection in the course of

ordinary lie detection examinations, research in other and more subtle methods should be encouraged.

2. In view of the limited objectivity with which the respiratory and plethysmographic response patterns can be visually evaluated (a) completely objective measurement methods (perhaps with computer aid) should be developed;
(b) new specific indices of deception should be structured and tested for validity.

3. In view of the large number of transducers and circuits available for use in lie detection, a systematic evaluation of these should be undertaken in order to determine which is the most reliable and most accurate within each physiological system.

4. In view of the limited objectivity with which the respiratory and plethysmographic reactions can be evaluated, and in view of the failure of this investigation to obtain stable discriminant functions among the various raters, the use of a computer as a decision machine in the lie detection process cannot be considered feasible at this time. On the other hand, development of computer type programs to objectify the analysis and measurement of complicated physiological patterns, such as are found in lie detection records, should be encouraged.

REFERENCES

Block, J. D., Rouke, F. L., Salpeter, M. M., Tobach, E., Kubis, J. F., & Welch, L., An attempt at reversal of the truth-lie relationship as measured by the psychogalvanic response. J. Psychol. 1952, 34, 55-66.

Dana, H. J., It is time to improve the polygraph: A progress report on polygraph research and development. In V. A. Leonard (Ed.) Academy Lectures on Lie Detection. Vol. II. Springfield, Ill.: Thomas, 1958, Pp. 84-94.

Dana, H. J., & Barnett, C. C., The emotional stress meter. In V. A. Leonard (Ed.) Academy Lectures on Lie Detection. Vol. I. Springfield, Ill.: Thomas, 1957, Pp. 73-83.

Davis, R. C., Physiological responses as a means of evaluating information. In A. D. Biderman & H. Zimmer (Eds.) The manipulation of human behavior. New York: Wiley, 1961, Pp. 142-168.

Holmes, N. D., The degree of objectivity in chart interpretation. In V. A. Leonard (Ed.) Academy Lectures on Lie Detection. Vol. II. Springfield, Ill.: Thomas, 1958, Pp. 71-83.

Inbau, F. E., & Reid, J. E., Lie detection and criminal interrogation. (3rd ed.) Baltimore: Williams & Wilkins, 1953.

Joseph, C. N., Analysis of compensatory responses and irregularities in polygraph chart interpretation. In V. A. Leonard (Ed.) Academy Lectures on Lie Detection. Vol. I. Springfield, Ill.: Thomas, 1957, Pp. 93-99.

Kubis, J. F., Instrumental, chemical, and psychological aids in the interrogation of witnesses. J. soc. Issues., 1957, 13, 40-49.

Orear, J. A., New approach to inspection. Bull. Atomic Scientist. 1961, March, 107-110.

Thorndike, E. L., & Lorge, I., The teacher's word book of 30,000 words. New York: Teachers Coll., Columbia Univ., Bureau of Publications, 1944.

APPENDIX A

Simulated Theft Experiment

1. Instructions and Directions
2. Questionnaires
3. Rating Scales

INSTRUCTIONS AND DIRECTIONS

Simulated Theft Experiment

DIRECTIONS TO SUBJECTS

I. GENERAL RECRUITMENT

A. To potential subjects:

"We want you to take part in an important scientific study.

You are to regard the details of this research and your participation in it as a confidential matter. The seriousness with which we consider this work prompts us to ask you to sign a pledge to reveal to no one any of the details in the study, or your part in it.

We would like your complete cooperation. This means doing exactly as you are instructed. Be assured that there will be no trickery. Neither will there be impossible tasks for you to perform.

A government grant has made it possible for us to pay you $1.50 for the one session in which you will participate. The session will last between an hour and an hour and a half.

There will also be an opportunity for you to be eligible for a number of cash prizes, one being $50. This will depend on how well you comply with the directions you are given."

II. SPECIFIC INSTRUCTIONS BY SUPERVISOR (Theft Situation)

A. To the "innocent" suspect:

"Your task is to enter the examining room and answer truthfully any questions which will be asked.

There has been a theft of some money from a religious pamphlet rack. You will be questioned about possible participation in it. For example, if you were asked "Did you steal the money or serve as an accomplice in this robbery?" you would, naturally, answer (NO). Since you know nothing about it and are completely innocent, there is nothing to worry about.

It is imperative that you do not discuss the nature of the questions or the research itself with anyone - not any of your friends nor your family.

Also, don't ask the examiner any questions because his time is limited. I'll answer your questions after the test.

How well you comply with these directions will determine whether you are eligible for cash prizes.

Are there any questions?

B. To the "thief" and "accomplice":

[To Both] "Please choose one of these lots. When you open it you will know your role in this research. Sign your name on the lot. (Supervisor notes name on lot.) Now put it in your pants pocket.

There is a pamphlet rack in the basement of Keating Hall, in the corridor leading to the men's room. (Point out corner of building.)

[To Thief] You, (Point to Thief.) must approach the rack and rob the coinbox of all its contents. You may approach in any manner you desire, but do not get caught.

While you are robbing the pamphlet rack you will have the help of your accomplice. (Point to him.)

[To Lookout] Your job is to act as his lookout - to warn him of approaching students or faculty. You will warn him in a manner which you both agree on, if danger is imminent.

You, however, are not to take the money. Neither are you to be close enough to him (Point to Thief.) to know what has been taken.

[To Thief] And neither should you tell him.

Now, when you have completed the robbery, signal the lookout to proceed upstairs.

[To Lookout] After he signals, you are to come upstairs alone, immediately. Return to Room _____.

[To Thief] When he (Point to Lookout.) has gone and you are a safe distance from the rack, examine what you have taken. Count the money. Put everything you have taken from the coinbox into your pants pocket (Supervisor indicates.) Return to Room _____

as soon as possible, but don't run.

[To Both] When you fellows return, you will be taken to the examining room where you will be asked questions about the robbery of the coinbox.

The examiner knows that the coinbox will have been robbed but he does not know who did it.

Now neither of you are to admit any guilt or participation in this matter. For example, if either of you are asked "Did you steal the money or serve as an accomplice in this robbery" you are to say (NO).

Is that clear?

Remember, deny any participation in or knowledge of the robbery. Answer NO to all questions about the robbery.

If you comply with the directions that you have been given, you will be eligible for cash prizes.

Any questions?"

III. INSTRUCTIONS BY EXAMINER DURING LIE DETECTOR TEST

A. Pre-Test Period

"As you know this is a lie detection machine similar to the ones used in police work. The innocent person has nothing to fear, since the procedure clearly proves his innocence. The matter is quite different for the guilty person. These machines have an excellent record in identifying the criminal. And the criminal has a difficult job concealing his guilt. (PAUSE)

I'd like you to sit in this chair so that I can attach you to the machine. (STRAP IN)

Are you in a comfortable position? Please don't make any movements. They only spoil the records. Your cooperation in avoiding movements is requested so that we don't have to waste any time retesting.

B. Testing Period

1. ASSOCIATION TEST

"I'm going to say a number of words. I want you to say the first word that comes to your mind when you hear each word that I say. READY?

apple; eagle; button; pencil; far-off; lookout;
roadside; warlike; coinbox; springtime; ocean.

(Retest)

Now I'm going to repeat the same words. Again, your task will be to give the first word that comes to your mind. Any questions?"

(Examiner repeats association words again.)

2. QUESTIONNAIRE A

"Now, I will ask a number of questions, some of which will be about the robbery. To all questions give only a YES or NO answer. Some questions will be repeated more than once.

There are no trick questions. All of them can be answered by a YES or by a NO. READY?"

(Questionnaire A is then administered.)

3. NUMBERS TEST

"This time I am going to say some numbers.

These refer to money, so that 78 means 78 cents and 212 means 2 dollars and 12 cents.

Now when I say a number, you are to say the number that follows the one I give. For example, if I say 78 you would say _____?"

84 99 114 124 134 139

4. QUESTIONNAIRE B

"Now I am going to ask you some more questions. Again, please answer only with a YES or NO. As before, some of the questions will be asked more than once."

QUESTIONNAIRES

Simulated Theft Experiment

Name _____ University _____
Last First Middle

Date of Birth _____ School _____
(within university)

Place of Birth _____
City State Country

Name of High School _____ Location _____

CIRCLE CORRECT ALTERNATIVES:

Sex . . . . . M F
Marital Status . . . . . Single Married
Year in College . . . . . Freshman Soph Junior Senior
No. of Brothers . . . . . 0 1 2 3 4 Other _____
No. of Sisters . . . . . 0 1 2 3 4 Other _____
Draft Status 1A 2A 2S 4A 4F

Mother Living? . . . . . Yes No
Father Living? . . . . . Yes No
Are you a U. S. citizen? . . . . . Yes No
Do you own a car? . . . . . Yes No
Were you ever fingerprinted? . . . . . Yes No
Are you now in the ROTC? . . . . . Yes No
Do you live on campus? . . . . . Yes No

Do you have a driver's license? . . . . . Yes No
Do you have any morning classes? . . . . . Yes No
Do you have any afternoon classes? . . . . . Yes No
Were you ever a patient in a hospital? . . . . . Yes No
Have you ever had a course in psychology? . . . . . Yes No
Have you ever heard of the Fordham Pathometer? . . . . Yes No

Pledge: I, _____, solemnly promise not to
(print your name in full)
reveal any of the details of this experiment to anyone, nor to discuss it with any other participant for a period of one year from this date.

Signed: _____

_____
(Date)

QUESTIONNAIRE A

(Second Student Experiment)

	1.	Is your name _____?
	2.	Did you eat breakfast this morning?
	3.	Is your draft status 1A?
I.	4.	Were you an accomplice to the thief?
	5.	Do you live in Brooklyn?
	6.	Are you married?
II.	7.	Did you take the money from the coinbox?
	8.	Do you have any afternoon classes?
	9.	Have you ever been fingerprinted?
III.	10.	Do you have the coinbox money with you?
	11.	Are you a college student?
	12.	Is your draft status 1A?
II.	13.	Did you take the money from the coinbox?
	14.	Do you live in Brooklyn?
	15.	Are you married?
III.	16.	Do you have the coinbox money with you?
	17.	Do you have any afternoon classes?
	18.	Have you ever been fingerprinted?
I.	19.	Were you an accomplice to the thief?
	20.	Are you a college student?
	21.	Is your draft status 1A?
III.	22.	Do you have the coinbox money with you?
	23.	Do you live in Brooklyn?
	24.	Have you ever been fingerprinted?
II.	25.	Did you take the money from the coinbox?
	26.	Are you a college student?
	27.	Are you married?
I.	28.	Were you an accomplice to the thief?
	29.	Do you have any afternoon classes?
	30.	Is your name _____?

QUESTIONNAIRE B

(Second Student Experiment)

	1.	Are you an American citizen?
	2.	Do you have any morning classes?
	3.	Is your mother living?
I.	4.	Do you know how much money was in the coinbox?
	5.	Do you live on campus?
	6.	Are you now in a Reserve Officer Training Program?
II.	7.	Did you act as a lookout for the thief?
	8.	Were you born in the Bronx?
	9.	Were you ever operated on for appendicitis?
III.	10.	Do you have the coinbox money with you?
	11.	Did you graduate from high school?
	12.	Is your mother living?
II.	13.	Did you act as a lookout for the thief?
	14.	Do you live on campus?
	15.	Were you ever operated on for appendicitis?
I.	16.	Do you know how much money was in the coinbox?
	17.	Did you graduate from high school?
	18.	Are you now in a Reserve Officer Training Program?
III.	19.	Do you have the coinbox money with you?
	20.	Were you born in the Bronx?
	21.	Were you ever operated on for appendicitis?
II.	22.	Did you act as a lookout for the thief?
	23.	Did you graduate from high school?
	24.	Are you now in a Reserve Officer Training Program?
I.	25.	Do you know how much money was in the coinbox?
	26.	Were you born in the Bronx?
	27.	Is your mother living?
III.	28.	Do you have the coinbox money with you?
	29.	Do you live on campus?
	30.	Do you have any brothers?
	31.	Before today, did anyone tell you the details of this theft?
	32.	Are you an American citizen?

RATING SCALES

Simulated Theft Experiment

Name: _____ Date: _____ Role: Innocent

By checking the appropriate place on the rating scales, describe your emotional reactions and feelings.

I. While waiting (just before taking the Lie Detector Test).

Relaxed	____ very relaxed	____ moderately relaxed	____ more relaxed than tense	____ more tense than relaxed	____ moderately tense	____ very tense	Tense
Calm	____ very calm	____ moderately calm	____ more calm than excited	____ more excited than calm	____ moderately excited	____ very excited	Excited
Disinterested & detached from procedure	____	____	____	____	____	____	Interested & involved in my role

Any other comments (feelings or unusual happenings):

_____

II. During the interrogation (i.e. during the Lie Detector Test).

Relaxed	____	____	____	____	____	____	Tense
Calm	____	____	____	____	____	____	Excited
Disinterested & detached from procedure	____	____	____	____	____	____	Interested & involved in my role

Any other comments (feelings or unusual happenings):

_____

III. During the Lie Detector Test, did you feel that your records may have indicated guilty reactions? (Refer only to questions about the crime.)

___ ALL of my responses probably indicated guilty reactions.
___ MOST of my responses probably indicated guilty reactions.
___ MORE THAN HALF of my responses probably indicated guilty reactions.
___ LESS THAN HALF of my responses probably indicated guilty reactions.
___ VERY FEW of my responses probably indicated guilty reactions.
___ NONE of my responses probably indicated guilty reactions.

Name: _____ Date: _____ Role: Lookout

By checking the appropriate place on the rating scales, describe your emotional reactions and feelings.

I. While you were acting as lookout.

Relaxed _____ _____ _____ _____ _____ _____ Tense
very relaxed | moderately relaxed | more relaxed than tense | more tense than relaxed | moderately tense | very tense

Calm _____ _____ _____ _____ _____ _____ Excited
very calm | moderately calm | more calm than excited | more excited than calm | moderately excited | very excited

Disinterested & detached from procedure _____ _____ _____ _____ _____ _____ Interested & involved in my role

Any other comments (feelings or unusual happenings):

_____

II. While waiting (just before taking the Lie Detector Test).

Relaxed _____ _____ _____ _____ _____ _____ Tense

Calm _____ _____ _____ _____ _____ _____ Excited

Disinterested & detached from procedure _____ _____ _____ _____ _____ _____ Interested & involved in my role

Any other comments (feelings or unusual happenings):

_____

III. During the interrogation (i.e. during the Lie Detector Test).

Relaxed _____ _____ _____ _____ _____ _____ Tense

Calm _____ _____ _____ _____ _____ _____ Excited

Disinterested & detached from procedure _____ _____ _____ _____ _____ _____ Interested & involved in my role

Any other comments (feelings or unusual happenings):

_____

Rating Scales: Lookout (cont'd)

IV. During the Lie Detector Test, did you feel that your recorded reactions gave you away? (Refer only to questions about the crime.)

___ ALL of my reactions probably showed that I was lying.
___ MOST of my reactions probably showed that I was lying.
___ MORE THAN HALF of my reactions probably showed that I was lying.
___ LESS THAN HALF of my reactions probably showed that I was lying.
___ VERY FEW of my reactions probably showed that I was lying.
___ NONE of my reactions probably showed that I was lying.

V. While you were acting as lookout, there were (other than the thief)

___ more than three people nearby or passing by.
___ one to three people nearby or passing by.
___ no people nearby or passing by.

Name: _____ Date: _____ Role: Thief

By checking the appropriate place on the rating scales, describe your emotional reactions and feelings.

I. While you were taking and pocketing the money.

Relaxed ____ _____ _____ _____ _____ ____ Tense
very relaxed | moderately relaxed | more relaxed than tense | more tense than relaxed | moderately tense | very excited

Calm ____ _____ _____ _____ _____ ____ Excited
very calm | moderately calm | more calm than excited | more excited than calm | moderately excited | very excited

Disinterested & detached from procedure _____ _____ _____ _____ _____ _____ Interested & involved in my role

Any other comments (feelings or unusual happenings):

_____

II. While waiting (just before taking the Lie Detector Test).

Relaxed _____ _____ _____ _____ _____ _____ Tense

Calm _____ _____ _____ _____ _____ _____ Excited

Disinterested & detached from procedure _____ _____ _____ _____ _____ _____ Interested & involved in my role

Any other comments (feelings or unusual happenings):

_____

III. During the interrogation (i.e. during the Lie Detector Test).

Relaxed _____ _____ _____ _____ _____ _____ Tense

Calm _____ _____ _____ _____ _____ _____ Excited

Disinterested & detached from procedure _____ _____ _____ _____ _____ _____ Interested & involved in my role

Any other comments (feelings or unusual happenings):

_____

Rating Scales: Thief (cont'd)

IV. During the Lie Detector Test, did you feel that your recorded reactions gave you away? (Refer only to questions about the crime.)

___ ALL of my reactions probably showed that I was lying.
___ MOST of my reactions probably showed that I was lying.
___ MORE THAN HALF of my reactions probably showed that I was lying.
___ LESS THAN HALF of my reactions probably showed that I was lying.
___ VERY FEW of my reactions probably showed that I was lying.
___ NONE of my reactions probably showed that I was lying.

V. While you were taking and pocketing the money, there were (other than the lookout)

___ more than three people nearby or passing by.
___ one to three people nearby or passing by.
___ no people nearby or passing by.

APPENDIX B

Denial-of-Actual-Crime Experiment

1. Instructions

2. Questionnaires

INSTRUCTIONS TO REHABILITATION SUBJECTS

(Supervisors)

"We are testing a new machine here at the Clinic and we need some people to help us. I am sure that this would be an interesting experience for you, and it could give you some knowledge about scientific research. Also we'll be able to pay you for your time.

We want to check how well the machine reveals bodily reactions such as voice, breathing, and perspiration, when people are asked embarrassing or emotional questions. Not all of the questions that may be asked are embarrassing, as for example - Is your name _____? An example of a somewhat emotional question might be: Were you ever in an automobile accident? None of the questions are about sexual matters and there are no trick questions.

If you agree to take part in this work, we don't want to know any trouble that you may have had with the law. So if any question pertaining to such matters happens to be asked, you should answer the question by saying NO. Is that clear? Remember just simply answer NO.

You can be sure that no information is transmitted from the clinic files to the examiner. He knows nothing about you personally. All we are interested in seeing is how the machine records bodily changes to emotional or embarrassing questions.

If you are interested, we will schedule an appointment."

(Examiners)

"This is our equipment designed to measure bodily reactions of people.

I'm going to ask a number of questions and you are to answer each one by saying either YES or NO. Remember answer each question by a YES or NO answer.

Are you ready?"

QUESTIONNAIRE I

(Rehabilitation Subjects)

1. Is your name _____?
2. Were you born in the USA?
3. Did you eat breakfast this morning?
4. Is your mother living?
*5.
6. Do you ever ride the subway to work?
7. Have you ever been in an auto accident?
8.
9. Did you go to public school?
10. Is your draft status 1A?
11.
12. Do you ever travel by taxi?
13. Do you have any sisters?
14.
15. Do you live in an apartment?
16. Have you ever been in an auto accident?
17.
18. Do you ever travel by taxi?
19. Do you have any sisters?
20.
21. Do you ever ride the subway to work?
22. Is your mother living?
23.
24. Did you go to public school?
25. Is your draft status 1A?
26.
27. Do you live in an apartment?
28. Do you have any sisters?
29.
30. Do you ever travel by taxi?
31. Have you ever been in an auto accident?
32.
33. Do you live in an apartment?
34. Is your draft status 1A?
35.
36. Do you ever ride the subway to work?
37. Is your mother living?
38.
39. Did you go to public school?
40. Is your name _____?

*Two questions specific to each Rehabilitation patient's criminal record were entered into the blank spaces. Two other questions, pertaining to crimes in which the patients were not involved, were used as controls.

QUESTIONNAIRE II

(Rehabilitation Subjects)

1. Are you a United States Citizen?
2. Did you eat breakfast this morning?
3. Did you go to work today?
4. Were you ever held back in school?
*5.
6. Were you born in New York City?
7. Do you have a driver's license?
8.
9. Did you read today's paper?
10. Have you been in the armed forces?
11.
12. Do you live in Brooklyn?
13. Have you ever spent the night in a hospital?
14.
15. Are you righthanded?
16. Do you have a driver's license?
17.
18. Were you born in New York City?
19. Have you ever spent the night in a hospital?
20.
21. Do you live in Brooklyn?
22. Have you been in the armed forces?
23.
24. Are you righthanded?
25. Were you ever held back in school?
26.
27. Did you read today's paper?
28. Do you have a driver's license?
29.
30. Are you righthanded?
31. Have you been in the armed forces?
32.
33. Were you born in New York City?
34. Have you ever spent the night in a hospital?
35.
36. Did you read today's paper?
37. Were you ever held back in school?
38.
39. Do you live in Brooklyn?
40. Are you a United States Citizen?

*Two questions specific to each Rehabilitation patient's criminal record were entered into the blank spaces. Two other questions, pertaining to crimes in which the patients were not involved, were used as controls.

APPENDIX C

Denial-of-Classified-Information Experiment

1. Instructions to Subjects

INSTRUCTIONS: DENIAL-OF-"CLASSIFIED"-KNOWLEDGE EXPERIMENT

A. Points Covered By Administrators In Instructions To Their Personnel

1. This research is important and valuable.
2. We are asking for volunteers.
3. There will be one session. On Tuesday and Wednesday of next week, we will try to retest as many of you as possible.
4. Essentially this is to see how well people can withold important information from individuals who have no need to know.
5. The questioning will involve classified material discussed on Monday.
6. The examiners may ask you about several details relating to the matter discussed. But since they have no need to know, you are justified and obligated to answer NO to questions pertaining to the case.
7. Obviously, you must deny knowledge of any classified material.
8. Your identity will be safeguarded. We won't know about your performance. Thus, how you perform in the research will not influence your status here.

B. Supervisor To Subjects

"We want to take part in an important scientific study.

You are to regard the details of this research and your participation in it as a confidential matter. Because of the importance of this work we request that you reveal to no one any details of the study, or your part in it.

To assure anonymity we will now give you a code number. It is the number on the tag which you are to keep in your wallet until Thursday (Friday). This code number can in no way reveal your personal identity.

Furthermore, your performance in this research will in no way influence your status here.

You will be asked a number of questions, some of which may pertain to classified material.

Please answer all questions by YES or NO.

Are there any questions."

C. Examiner To Subjects

"This is a lie detection machine similar to the ones used in police work.

These machines have an excellent record in identifying criminals. And the criminal has a difficult job concealing his guilt.

I'd like you to sit in this chair so that I can attach you to the machine.

I will ask a number of questions. To all questions give only a YES or NO answer. There are no trick questions. All of them can be answered by a YES or by a NO.

Are you in a comfortable position?

Please don't make any movements. They only spoil the records. Your cooperation in avoiding movements is requested so that we don't have to waste any time retesting.

APPENDIX D

Counter-Measures Experiment

1. Instructions to Subjects

EXAMINER'S INSTRUCTIONS TO SUBJECTS

(Counter-measures Experiment)

"There are four different parts to this session. Deliberate violation of instructions will disqualify you from the prize money."

PART A (Contol Test)

"Now for the first part.

You are to select one of these slips which will have a number on it. Read the number (to yourself) and keep it in mind. Then put the slip into your pocket.

I'm going to try to find out the number on the slip. Each time I call out a number, even though it be the number you have chosen, you are to say NO. In other words you are to say NO to every number I say. Is that clear?

Now, for the important instructions. Although you are to deny or lie about your number, do not try to use any deliberate means to 'beat the machine.' In other words, I don't want you to move, wiggle, cough or breathe hard so as to 'foul' up this procedure. I merely want to find out what happens in lying under ordinary conditions.

Remember, I want you to lie about your number. However, if you make any deliberate attempts to contaminate the record, and these are not hard to spot, you will be disqualified.

Do you remember the number on the slip? Fine! Now I will begin with the test.

Ready."

(Number series is presented.)

PART B (Complete Relaxation)

"Now for the second part.

Here we will ask you to use a special method to beat the machine. This is the RELAXATION or YOGA method. You are to relax completely: every muscle--in the arms, the legs, face and neck.

You are to try to get into an 'abstract' or 'far away' frame of mind. Just try to think of nothing. You'll hear everything I say but

EXAMINER'S INSTRUCTIONS (cont'd)

you'll consider it as a voice far off in the distance. Being relaxed, you will not be affected or excited when hearing the numbers I say.

Let us practice a bit and see if you can get into this relaxed 'far away' frame of mind. Try relaxing your muscles from head to toe. (Examiner encourages the subject to relax completely.)

While you are in this relaxed and abstracted state I'm going to ask you a few questions. You are to answer them by YES or NO in a relaxed manner. We want to check on how well the instrument indicates complete relaxation.

Questions

1. Are you feeling relaxed?
2. Are you a graduate student?
3. Were you born in Europe?
4. Do you live on Staten Island?

(If necessary the examiner repeats the last three questions until the electrodermal response is extinguished.)

O. K. That's fine.

Take one of the slips. Look at the number, remember it, and put it into your pocket. Again, I'll say a set of numbers and you are to answer NO to each number. But this time you will be answering while in this completely RELAXED state.

Now just relax and when you are completely relaxed say "NOW" and I'll begin with the numbers."

(The second list of numbers is used.)

PART C (Muscle Tension)

"Now for the third part.

This time you'll try to beat the lie detector by using the MUSCLE TENSION procedure. This requires tensing muscles which are not visible and which have no immediate connection with the lie detection equipment, as for example, the muscles of the toes, the calf, or the internal stomach muscles. However, movements of the hand, arm or head are easily observed and readily detected by the machine.

EXAMINER'S INSTRUCTIONS (cont'd)

First, you are to practice tensing the toe muscles. Press them against the floor. Try this until you can do it rather easily and without moving the shoes.

Now I'm going to ask you the same questions again. You are to tense your toe muscles on one of the questions.

Questions

1. Are you ready?
2. Are you a graduate student?
3. Were you born in Europe?
4. Do you live on Staten Island?

(Check what question was used in the TENSION state. If there are no clear indications on the chart, repeat the practice procedure.)

O.K. That's fine.

Now we are ready for the test. Select a slip. Note the number on it and put it in your pocket.

Your job is to induce tension not on the number you have selected, but on a different number. In other words, you'll try to shift the tension reaction from the number on the slip to another number in the series.

For example, after the first call of the number on your slip, you are to tense your muscles on either of the two numbers that will follow it. Thereafter you are to stick with that number and you are to tense your muscles only to that same number whenever it is asked. For example: (Here the examiner, with a hypothetical set of numbers, demonstrates the procedure to be followed by the subject.)

Is that clear? Remember to tense your muscles always on the same number.

Incidentally, are you sure of the number on your slip? Fine. Then we'll begin."

(The third number series is used.)

PART D (Exciting Imagery)

"Now for the final part.

To try to beat the lie detector, you will use the EXCITING IMAGERY technique.

EXAMINER'S INSTRUCTIONS (cont'd)

Recall to mind some incident in your life, or some thought or image that you know can get you excited or upset. Do you have something exciting in mind for the practice session? Fine. You are to attach this EXCITING IMAGE to one of the questions I will ask. These will be the same questions asked before. Remember, use the EXCITING IMAGE on only one question, but not the one used in MUSCLE TENSION.

Questions

1. Are you ready?
2. Are you a graduate student?
3. Were you born in Europe?
4. Do you live on Staten Island?

(Check the question used for EXCITING IMAGERY. If clear results are not apparent, repeat the practice session.)

O. K. That's fine.

Now select a slip. Read it, make a mental note of it, and put it into your pocket.

The best procedure in the EXCITING IMAGERY technique is to attach the exciting image to a number other than the one you have selected. For example, after you say NO to the number on the slip (i. e. the very first time you lie to that number) you are to arouse this exciting image on either of the two numbers that follow the lie. You are to stick to that number and to attach the exciting image to that same number thereafter.

(Examiner demonstrates the procedure again with a hypothetical set of numbers.)

Is that clear?

Do you remember the number on your slip? Fine! Then we'll begin."

(The fourth series of numbers are used.)

After the experiment

"Please do not talk with anyone about the procedures used in this experiment. If you do, you will forfeit your right to one of the prizes."

APPENDIX E

Recording Forms

Experiment # _____ Rater _____
Examiner _____ Date _____

(++) High Signif. (+) Signif. (?) Doubt Signif. (-) Non-Sig. (0) Non-interp.

RECORD #1		Subject 1			Subject 2			Subject 3		
		Resp	Pleth	PGR	Resp	Pleth	PGR	Resp	Pleth	PGR
button										
lookout										
coin box										
Accomplice	(4)									
	(19)									
	(28)									
Take money	(7)									
	(13)									
	(25)									
Money with	(10)									
	(16)									
	(22)									
button										
lookout										
coin box										

RECORD #2		Resp	Pleth	PGR	Resp	Pleth	PGR	Resp	Pleth	PGR
Numbers										
How much	(4)									
	(16)									
	(25)									
Lookout	(7)									
	(13)									
	(22)									
Money with	(10)									
	(19)									
	(28)									

DIAGNOSIS ( ) ( ) ( )

REMARKS

RATER CONFIDENCE SCALE

Experiment #_____ Rater _____
Examiner _____ Date _____

CONFIDENCE In own "Diagnosis"	Subject 1 Diagnosis ( )	Subject 2 Diagnosis ( )	Subject 3 Diagnosis ( )
Almost certain			
Very confident			
More confident than doubtful			
More doubtful than confident			
Very doubtful			
No confidence			

REMARKS:

RECORDS Easy to Interpret?	Resp	Pleth	PGR	Resp	Pleth	PGR	Resp	Pleth	PGR
Very easy									
Easy									
Difficult									
Very difficult									

REMARKS:

AIDS In Diagnosis	Ass'ns	Nos.	Ass'ns	Nos.	Assn's	Nos.
Most helpful						
Moderately helpful						
Of minor help						
Not helpful						

REMARKS:

Did Subject's Behavior
Influence your Diagnosis?
Comment on this if it did.

APPENDIX F

Photographs

1. Instrument and Attachments
2. Specimen Records

FIGURE 1

Photograph of the control and recording panels of the Grass Polygraph. The physiological reactions, respiration, blood pressure, and PGR, were calibrated and controlled by the upper two and bottom control panels, respectively. In the lower part of the photograph, the recording pens and a sample record are in evidence.

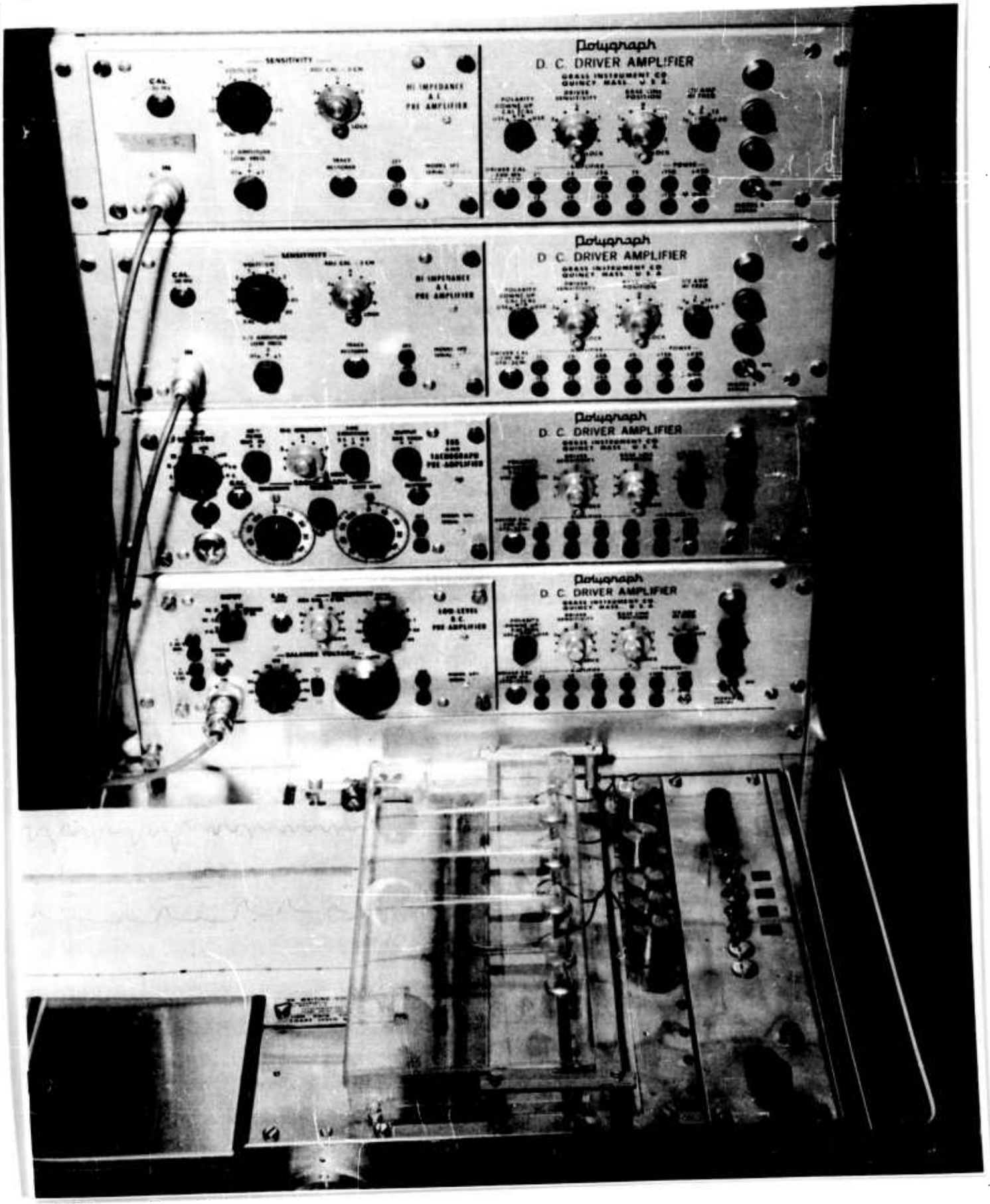
Polygraph
D. C. DRIVER AMPLIFIER
Polygraph
D. C. DRIVER AMPLIFIER
Polygraph
D. C. DRIVER AMPLIFIER
Polygraph
D. C. DRIVER AMPLIFIER

FIGURE 2

Photograph of a subject with the attachments necessary to record the three physiological measures. The rubber bellows around the chest provided the respiration record, while the metal finger oncometer on the middle finger of the subject's right hand gave rise to the plethysmographic record. The psychogalvanic response was obtained from palmar electrodes which were attached to the hands of the subject by means of the broad rubber bands. Hidden from view are two transducers placed on platforms attached to the rear legs of the chair. Also employed (but not pictured above) was a Lavelier microphone which hung from the subject's neck, resting on his chest just above the bellows.

FIGURE 3

Photograph of a sample testing situation, with the experimenter in the foreground and the subject at the right. In order to depict the attachments of the subject, including the Lavelier microphone, the subject was positioned to enable a side view. However, in actual experimentation, he was placed with his back to the experimenter.

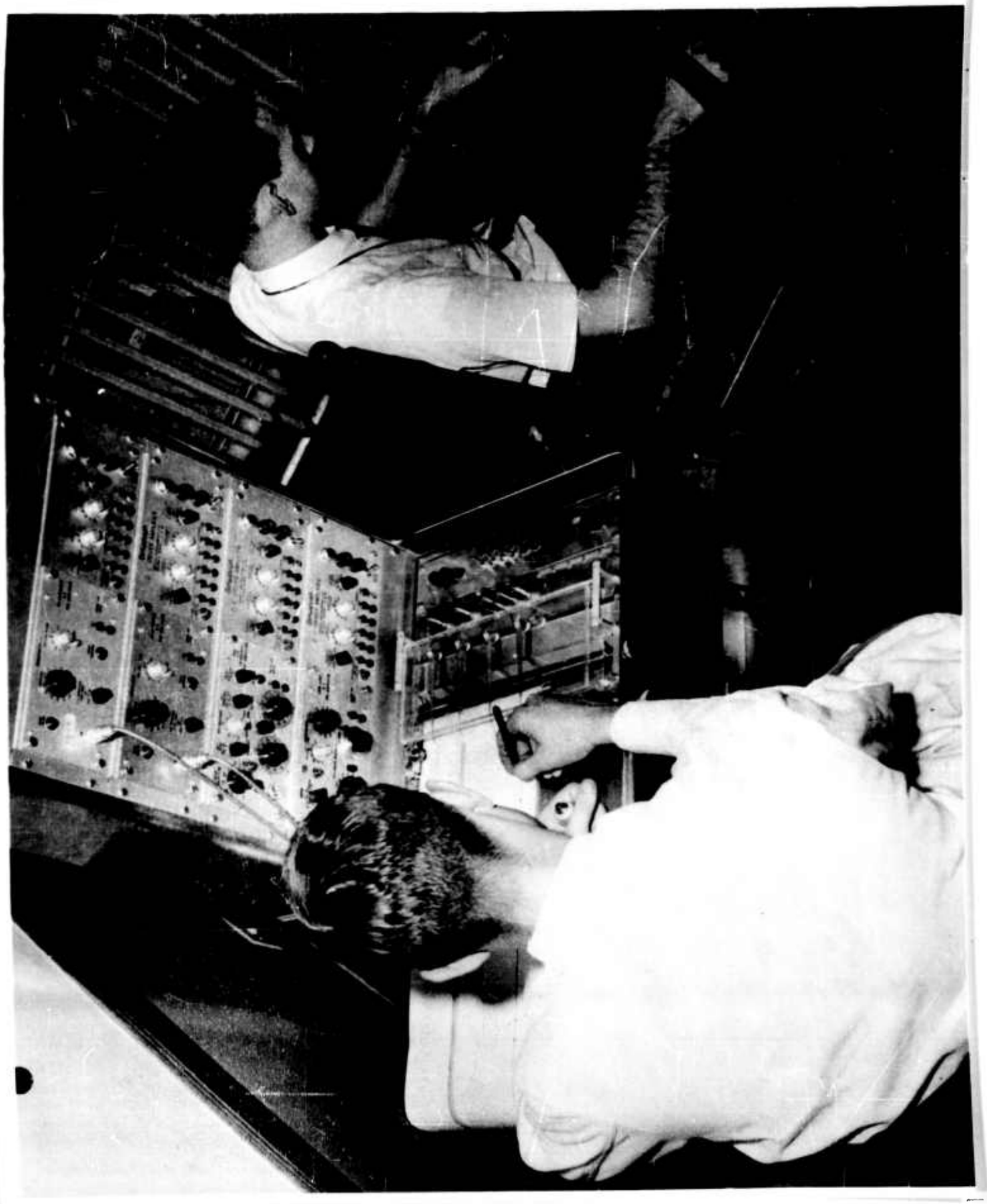

FIGURE 4

Photograph of a portion of an actual record indicating the respiration pattern (at top of chart), a relatively unresponsive plethysmographic pattern (middle recordings) contaminated by several movements (large spikes), and the psychogalvanic response (third recording) indicating very little reactivity except for a slow decrease in palmer resistance (rising curve). Resetting of the psychogalvanic response base line is noted by the numbers 240. The marker and timing pen produced the recording at the bottom of the chart where the large blocks of ink indicate the points at which stimuli (association words) were applied.

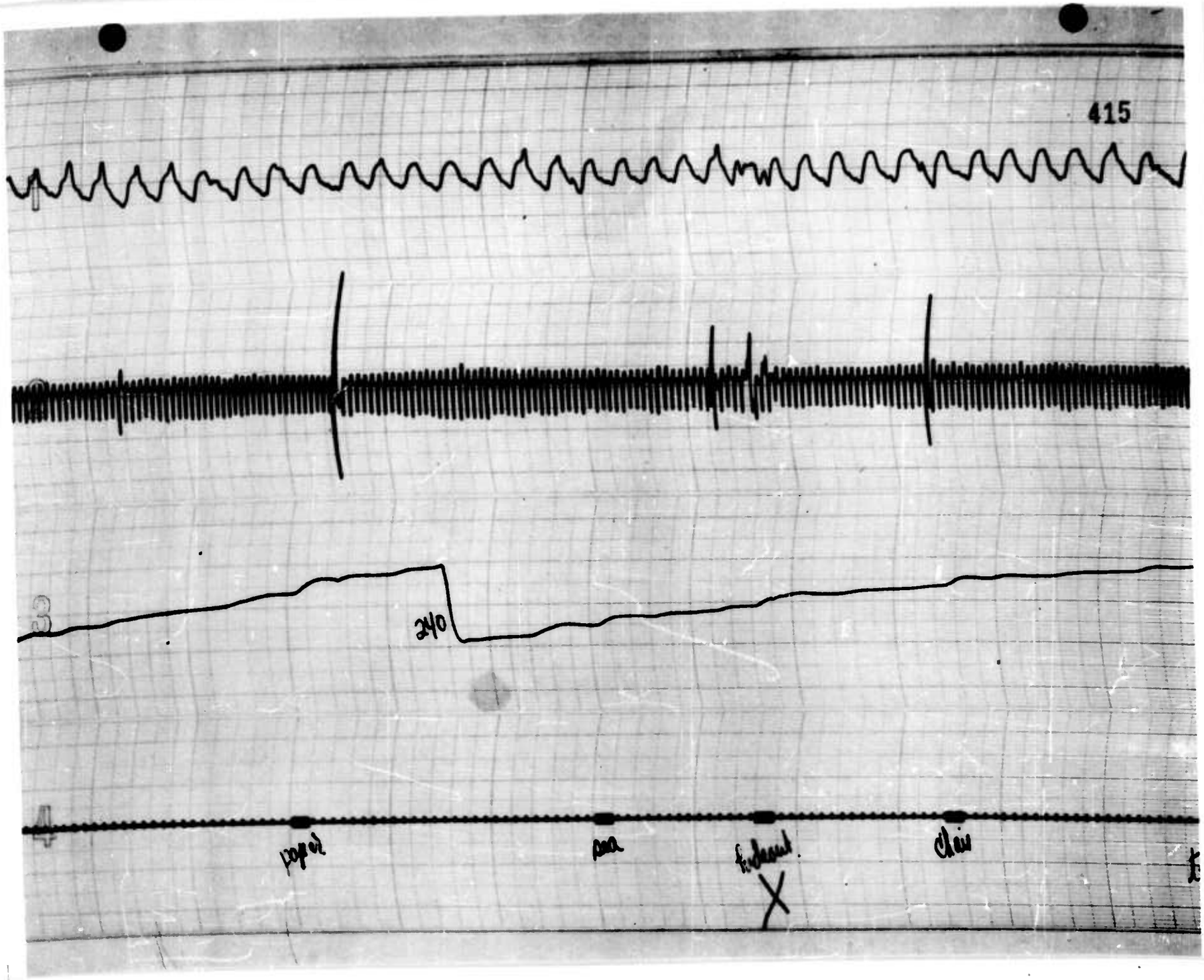
415
3
240
4
paper
pen
chart

FIGURE 5

Photograph of an actual record clearly indicating "involvement" on the part of the subject. Significant responses are evident in the plethysmographic record (at I, II, III) and in the psychogalvanic record (at I and II).

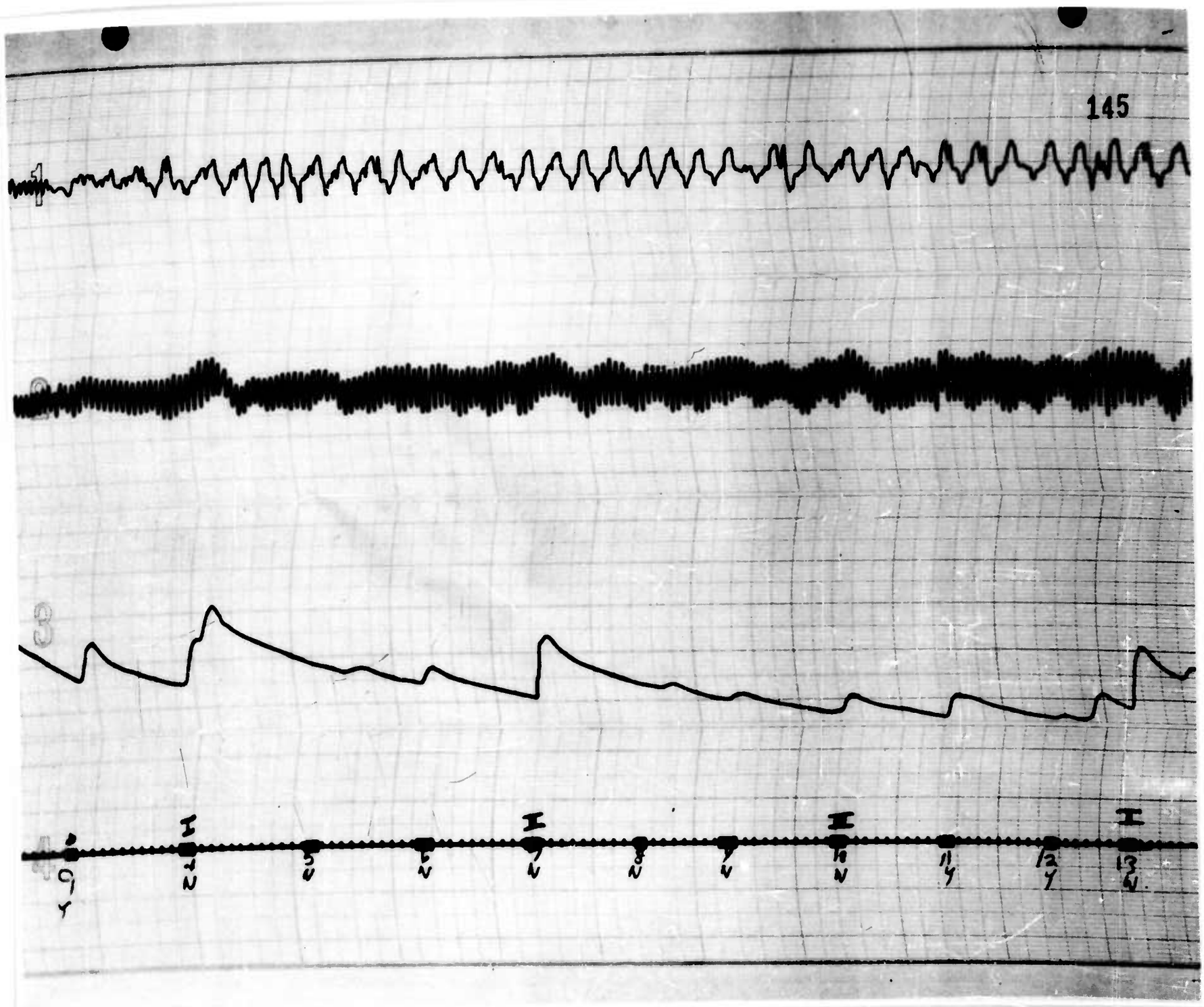
1
2
3
4
I
I
I
I
0 Y
1 N
2 Y
3 N
4 Y
5 N
6 Y
7 N
8 N
9 Y
10 N
11 Y
12 Y
13 N

FIGURE 6

Photograph of an actual record illustrating the effects of gross bodily movements (M) on the physiological recordings. These movements distort all three responses, in many cases causing deflections beyond the recording range of the pen units.

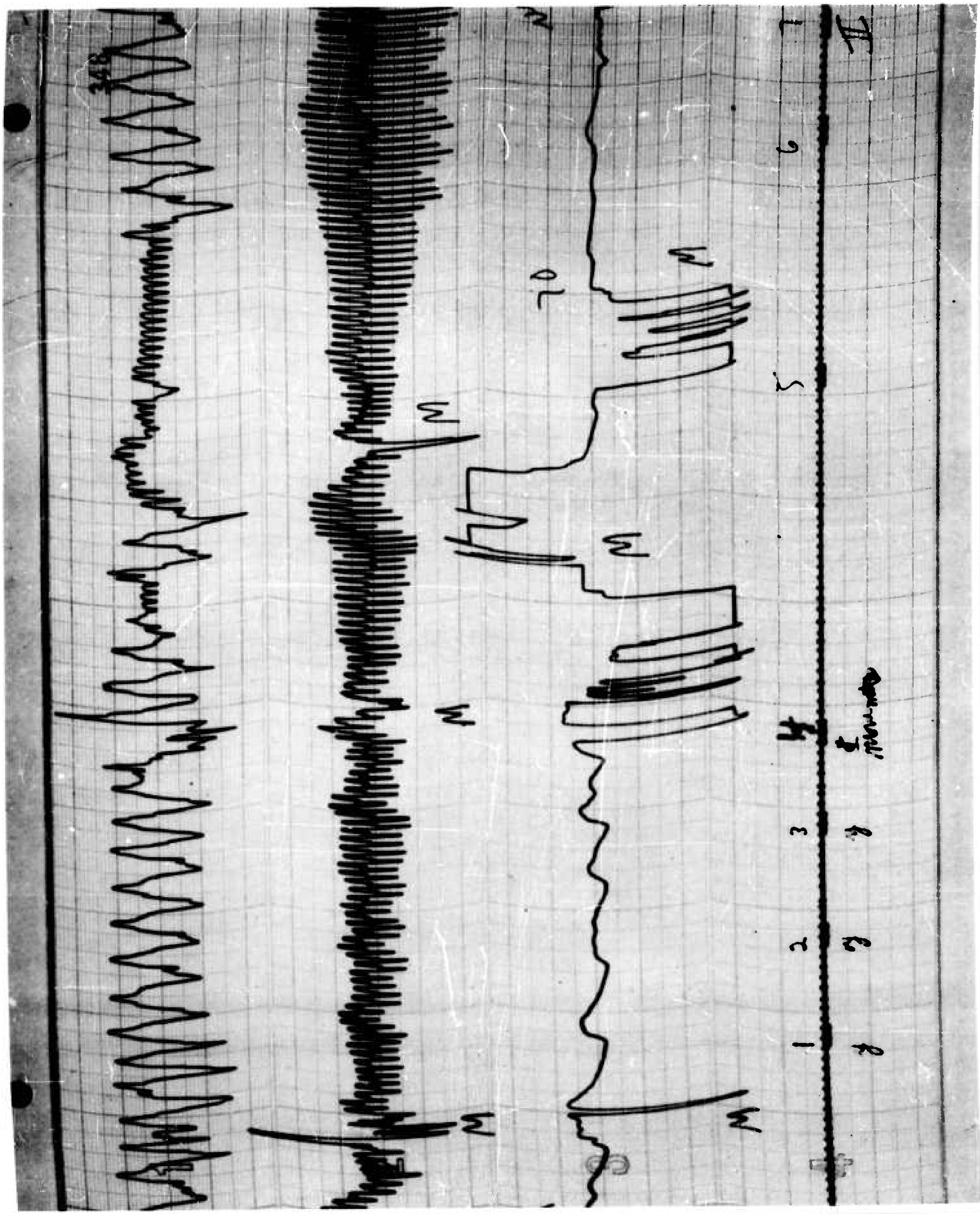

FIGURE 7

Photograph of an actual record obtained in the countermeasures experiment when the subject was attempting to deceive the operator by using the muscle tension technique. The number on which the subject was lying was 27, while muscle tension was used on number 22. When 27 was called for the first time, a significant psychogalvanic reaction along with movements in the plethysmographic record, were evident. The next number was 22, and the subject seemed to have held his breath in anticipation of his physical effort, which resulted in sizable psychogalvanic and plethysmographic reactions. When 22 was called again, the subject was very successful in producing significant psychogalvanic and plethysmographic responses.


262
B